

```
Db 301 TATAAATTTTAAAAATTTACTTTTAAATTTATAGTATTTTATTTTATTTTACTTAA 360
Qy 361 ttataaatctatctgaattttattgaattttttataattttatagttactataaaatacaaat 420
Db 361 TTATAAATTTATCTGAAATTTTATTTGAATTTTATTAATTTAGTACTTAAATAATACAAAT 420
Qy 421 atttttccctattctgaagaaaaattctatttttaaaaaaaattgatttttatagataaat 480
Db 421 ATTTTTCCTATTCTAAGAAAAATTCATTTTAAAAAAAATTTGATTTTATAGTATAAT 480
Qy 481 ttgtttgataattgaattgaattgaattgaattgaattgaattgaattgaattgaattgaatt 540
Db 481 TTGTTTGTATAAATGAATTAACTTTGAATTTGAAAGGGGAAACAAAATGAATAAATGCTTTGA 540
Qy 541 aaaaaattctgtattcatcagctatttatgcaacttcgcttgcatcaaatatttgcaattt 600
Db 541 AAAAAATCTTGATTCATCAGCTATTATGCAACTTCGCTGCTGCAATATTGCAATTT 600
Qy 601 gtgcagcaggttgtgacagacagaaatcaggttcaacttctgattctaaaccacaagcc 660
Db 601 GTTGACGAGGTTGTGACAGACAGAAATCAGGTTCAACTTCTGATTTCTAAACCACAAAGCC 660
Qy 661 gagacgtataaacaataaagtaagtaattgatttcttattogaaatgacactaaaccgatt 720
Db 661 GAGACGCTAAACATAAAGTAAGTAATGATTTCTATTCCGAATAGCAGCTAACCGATCCGGAT 720
Qy 721 aatctcgtatgaattagtgccaaaaagatatatttcttattcttattgtagtaaacagagcca 780
Db 721 AATCTCTGATGAATTAAGTGCCCAAAAGATATTATTTCTTATTGTGTGAACAGAGGCA 780
Qy 781 gcaacttcaacaattacaaaaaccaggtatgcacaaataactgactcactcagcaagct 840
Db 781 GCAACTTCAACAATTAACAAAAACAGGATGCACAAATAACTGACTCACTCAGCAAGCT 840
Qy 841 aatttaagccagcgcacaaaggaattattattgcccctgaaatggaatggaattgga 900
Db 841 AATTAAAGCCAGCGCCCAAAAGGATTTATTATTGCCCTGAAAATGGAATGGAGTTGGA 900
Qy 901 actgctgttaatacaattgctgataaaggaaattccgatttgcctatgactgaactaatt 960
Db 901 ACTGCTGTTAATCAATGCTGATAAGGAATTCGGATTTGCTGCTATGATCAGCTAAT 960
Qy 961 actggtatcgtataaataatgattggtatgttcttttattgataatgaaatggttgaatta 1020
Db 961 ACTGGATCTGATAAATATGATTGTTGTTCTTTTATGATAATGAAAAAGTTGTTGAATTA 1020
Qy 1021 caaggtcttcaacttgctgctgctgcttattaggaagaaagaaatggttcttttattgaatt 1080
Db 1021 CAAGGTCTTTCACPTGCTGCGGCTATTAGGAAAGAAAGATGCTGCTTTTGTATCAAT 1080
Qy 1081 gataaatgaatgaataatctaaatcacatatgcccccaagagacaatttctttttatata 1140
Db 1081 GATCAATGATGAATATCTAAATACATATGCCCCCAAGAGACAATTTCTTTTATATA 1140
Qy 1141 atcgcgggttcccaagatgataataatcccaatttttataatggtgcaatgaaagta 1200
Db 1141 ATCGCGGTTCCCAAGATGATAAATAATTCCTCAATATTTTATAATGTTGCAATGAAAGTA 1200
Qy 1201 cttaagaataatgaataattgcgcaaaataaaaataattgattttattctctgagcgaa 1260
Db 1201 CTTAAGAATTAATGAATAATTCGCAAAATAAATAAATTAATGATTATCTCTGAGGCGAA 1260
Qy 1261 aatgctgtttatgtccagagtgaaattatggaactccggttcaagaatcccaattcttt 1320
Db 1261 AATGCTGTTTATGTCAGAGTGAATTTATGGAATTCGCTGCTGCTGCTGCTGCTGCTT 1320
Qy 1321 ctacaattataaagatccagcaggttggttaataataatacaaatggttggttcaaaacca 1380
Db 1321 CTACAATTAACAAGATCCAGAGGTGTTAATAAATAAATAAAGCTGTTGTTTCAAAACCA 1380
Qy 1381 gcttctatttcaagattcttgcggcccaaatgaggaattgaggaatgcccgaacaagaatcacc 1440
Db 1381 ATGAAAAAATGCTTAGAAAAAATTTCTTATTATCATCAGCTATTATGCAATTCGCTT 1440
```

```
Db 1381 GCTTCTATTTCAAAGGATTTCTTGCCCAAAATGATGAATGCGCCGAACAAGCAATCAC 1440
Qy 1441 aaattaaactgaaggggttggataccacaaatacttggtaactgctcaagattataat 1500
Db 1441 AAATTAAACTTTGAAGGGTTGATACCCCAAAATCTTTGTAACTCGTCAAGATTATAAT 1500
Qy 1501 gataagcccaaaacttttcaaaagcggcgatcaaaatgacaaatttataaacctgat 1560
Db 1501 GATAAGCCCAAACTTTTATCAAGAGCGGATCAAAATATGACAATTTATTAACCTGAT 1560
Qy 1561 aaagttttgggaaagtctgctggaagtcttctgggttttaattgcaagaataataa 1620
Db 1561 AAAGTTTATAGGAAAGTCTGCTGTTGAAGTCTTCGGGTTTATTTATTCGAAGAAATAA 1620
Qy 1621 gcatctagatcagaagtcgaaacgaactaaagcaaaactacaaataatttcatttaa 1680
Db 1621 GCATCTAGATCAGAAGTCGAAACGAACTAAGCAAAACTACCAAAATTTTCATTATAA 1680
Qy 1681 tatgataatacaacatataaagtcacaggtacaaatattatacaaatatttagtaagcca 1740
Db 1681 TATGATAATCAACATATAAAGTACAAGTAAAAATATTATAACAATTTTAGTAAGTCCA 1740
Qy 1741 gtaattgttacaagaactaatgttgataatcctgagtcctaa 1782
Db 1741 GTAATTGTTACAAAGCTAATGTTGATAATCCTGATGCCTAA 1782

RESULT 2
Q22042 Q22042 standard; DNA; 1260 BP.
AC Q22042;
DT 02-NOV-1992 (first entry)
DE Sequence encoding surface antigen 46kd.
KW Swine pneumonia; epidemic; diagnosis; therapy; ss.
OS Mycoplasma hyopneumoniae.
FH Key
FT Location/Qualifiers
FT 1. 1260
FT /*tag= a
PN EP-475185-A.
PD 18-MAR-1992.
PF 27-AUG-1991; 114335.
PR 27-AUG-1990; JP-224945.
PA (NIFL-) NIPPON FLOUR MILLS.
PI Seto Y, Futo S, Mitsuse S, Matsuo K, Tsuna M;
DR WPI: 92-089874/12.
DR P-PSDB: R21829.
PT DNA and peptide of mycoplasma hyopneumoniae - useful for
PT diagnosis and treatment of swine mycoplasma pneumonia
PS Claim 2; Page 6 and pages 36-37; 45pp; English.
CC The inventors claim DNA encoding a surface antigen and primers used
CC in a method of diagnosing mycoplasma pneumonia of swine.
CC Mycoplasma cells were collected from the culture of M.h.p. lysed in a
CC buffer containing SDS, followed by the extraction of DNA and
CC purification thereof. The DNA of M.h.p is cleaved with a restriction
CC enzyme HindIII, the resulting fragment is inserted into a plasmid
CC pUC19 at the site cleaved with HindIII. This plasmid is transformed
CC into E. coli as a host cell, colony hybridization is performed by the
CC DNA probe method utilizing EcoRI fragments of pKUM1 and pKUM2 to
CC select the bacterial cells containing a plasmid (pUR126) carrying the
CC 46 kd antigenic gene. E. coli JM 109 which comprises the plasmid
CC pKUM1 or pKUM2 is deposited with FRI under accession No. FERM P-10318
CC or P-10319. The base sequence of the 46 kd antigen gene is given in
CC Q22042.
SQ Sequence 1260 BP; 476 A; 204 C; 212 G; 368 T;

Query Match 69.5%; Score 1239.2; DB 1; Length 1260;
Best Local Similarity 99.0%; Pred. No. 66-146;
Matches 1247; Conservative 0; Mismatches 13; Indels 0; Gaps 0.

Qy 523 atgaaaaaatgcttagaaaaaaattcttctgattctatcagctatttatgcaacttcgatt 1440
Db 1 ATGAAAAAATGCTTAGAAAAAATTTCTTATTATCATCAGCTATTATGCAATTCGCTT 60
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 1999, 11:53:40 ; Search time 47.89 Seconds
(without alignments)
9309.716 Million cell updates/sec

Title: US-08-913-430-1
Perfect score: 1782
Sequence: 1 atgaaaaaatgccactata.....ttgataacctgagcctaa 1782

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1777.2	99.7	1782	1 T38241	Mycoplasma 46-48 k
2	1239.2	69.5	1260	1 Q22042	Sequence encoding
3	835	46.9	843	1 Q05578	Sequence encoding
4	392.8	22.0	396	1 Q05576	Sequence encoding
5	356.8	20.0	360	1 Q05577	Sequence encoding
6	84	4.7	19124	1 T72882	Plasmodium var-7 g
7	82.8	4.6	9789	1 T41852	cDNA encoding Plas
8	81.6	4.5	5760	1 N50530	Sequence encoding
9	80	4.5	4590	1 N60472	Sequence encoding
10	77.6	4.4	2503	1 Q53480	PNP30 xylanase CD
11	77.4	4.3	1511	1 Q28302	AMEPV tk DNA, New
12	77.4	4.3	1511	1 Q65798	AMEPV thymidine-K1
13	77.4	4.3	1511	1 V14508	AMEPV entomopoxvir
14	76.6	4.3	1671	1 Q24134	50 kD subunit of S
15	76.6	4.3	2104	1 Q25273	Sequence encoding
16	75.2	4.2	1864	1 N71405	Sequence of ANS-1
17	75	4.2	7797	1 X33180	Cowpox virus bsr f
18	75	4.2	6644	1 X33181	Base sequence of t
19	75	4.2	7372	1 X33182	Base sequence of t
20	75	4.2	7996	1 X33184	Base sequence of t
21	74.4	4.2	8920	1 Q62924	Carbamoyl-phosphat
22	74.2	4.2	1864	1 Q78892	Aspergillus nidula
23	74	4.2	5852	1 Q11710	Dictyostelium plas
24	73	4.1	4590	1 N60472	Sequence encoding
25	72.2	4.1	3095	1 Q03875	Sequence encoding
26	71.8	4.0	605	1 T31530	Human 3' apolipop
27	71.2	4.0	998	1 X15422	Nicotiana tabacum
28	71	4.0	731	1 Q31693	Rsal restriction f
29	70.8	4.0	1942	1 Q12540	GSJA promoter elem
30	70.6	4.0	1782	1 T38241	Mycoplasma 46-48 k
31	70	3.9	605	1 T31530	Human 3' apolipop
32	70	3.9	6152	1 T78867	P. falciparum live
33	69.8	3.9	3101	1 Q02047	Sequence encoding
34	69.6	3.9	5852	1 Q11710	Dictyostelium plas
35	69.2	3.9	1864	1 N71405	Sequence of ANS-1
36	69.2	3.9	1939	1 Q34622	Cytosolic GSJA glu
37	69	3.9	3101	1 Q02047	Sequence encoding
38	68.8	3.9	32367	1 V35620	Human SHOX (short
39	68.8	3.9	15577	1 V35616	SHOX gene prelinin
40	68.4	3.8	3975	1 N81157	Malaria-specific g
41	68.4	3.8	3975	1 Q22999	SERP gene. Recombi
42	68.4	3.8	2104	1 Q25273	Sequence encoding
43	68.4	3.8	6152	1 T78867	P. falciparum live

44 68.4 3.8 110000 1 X20248-05 Continuation (6 of
c 45 68.2 3.8 731 1 Q31693 RsaI restriction f

ALIGNMENTS

RESULT 1
T38241
ID T38241 standard; DNA; 1782 BP.
AC T38241;
DT 19-DEC-1996 (first entry)
DE Mycoplasma 46-48 kda protective antigen gene.
KW Antigen; vaccine; mycoplasmal pneumonia; swine enzootic pneumonia;
diagnosis; antibody; ss.
OS Mycoplasma hyopneumoniae strain Beaufort.
FH Key Location/Qualifiers
FT 523..1782
FT /*tag= a
FT /transl_except= (730..732, aa:Trp)
FT /transl_except= (823..825, aa:Trp)
FT /transl_except= (1282..1284, aa:Trp)
FT /note= "IGA is a stop codon"
PN W09628472-Al.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149.
PF 16-MAR-1996; AU-001789.
PA (UTME) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
DR WPI: 96-433763/43.
DR P-PSDB; W01037.
PT Putative protective antigens against Mycoplasma - used for the
detection, prevention or treatment of Mycoplasma infections, esp. M.
hyopneumoniae in swine
PS Claim 14; Page 28-29; 43pp; English.
CC The gene (T38241) coding for a 48 kda putative protective antigen
(W01037) was obtd. from a Mycoplasma hyopneumoniae genomic library
by screening with a probe generated by PCR amplification (see also
T13814-16). The antigen had originally been isolated from M.
hyopneumoniae cells using antibody probes enriched with Mycoplasma-
specific antibodies. Other protective antigens were also identified
(see also W01024-31). Protective antigens and antibodies can be
used in vaccines for preventing or treating mycoplasmal infections,
partic. M. hyopneumoniae infections in swine. They can also be used
for diagnosis.
SQ Sequence 1782 BP; 682 A; 254 C; 259 G; 587 T;

Query Match 99.7%; Score 1777.2; DB 1; Length 1782;
Best Local Similarity 99.8%; Pred. No. 2.4e-212;
Matches 1779; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaaaaaatgccactataccagaggaaagcagtgatataataataataataataatcat 50
Db 1 ATGAAAAAATGCCACTATACCAGAGGAAGACGAGTATATAAATAATAATAATACAT 60
QY 61 ttcttcatttcgcccagcaatttttaagaatttagcattataaagaagtagacacaaagtta 120
Db 61 TTCTTCATTTCGCCGAGAAATTTTAAAGAAATAGTACATTAAAGTAAACAAAAGTTA 120
QY 121 ttaatgtaaacattagcgcaatcccttaagaaaaaattataattttttttttttttta 180
Db 121 TTAATGTAACATTAGCGCAATCCTTAGAAAAAATTAAGCTTTTATCTATTTTITA 180
QY 181 atcgaaatccaccaggcataaatctttgtcagttatttatcaagtcggtattttttcatt 240
Db 181 ATCGAAATCCACCAGGCAATAATCTTTGTCAAGTATTTATCAAGTCGGTATTTTTCATT 240
QY 241 atttctactaaatatttgaatttcattttccataatcctaaatttttttttttttttt 300
Db 241 ATTCTACTAAATATTATTGAAATTCATTTTCCATAATCTAAATTTTACATTTTTT 300
QY 301 tataacaatttttaaaaaattactcttatttagtagtatttttttttttttttagtctaaa 360


```

FT exon      801..995
FT FT        /*tag= a
FT exon      1199..4225
FT FT        /*tag= b
PN WO8601802-A.
PD 27-MAR-1986.
PF 11-SEP-1985; 006960.
PR 11-SEP-1984; AU-007067.
PR 11-SEP-1984; AU-007066.
PR 10-SEP-1985; AU-047326.
PA (HALL-) HALL INST MED RES.
PI Kemp DJ, Anders R, Coppel RL, Brown G, Saint RB, Cowman AF;
DR WPI; 86-094065/14.
DR P-PSDB; P60569.
DR DNA coding for Plasmodium falciparum antigens - expressing
PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P
PT falciparum.
PS Claim 4; Fig 1; 55pp; English.
CC The inventors claim a novel DNA molecule which comprises a
CC nucleotide sequence corresp. to all or a portion of the base
CC sequence coding RESA (N60473) or FIRA (N60473). RESA and FIRA have
CC antigenicity suitable for providing protective immunity against
CC Plasmodium falciparum malarial infections.
CC Sequence 4590 BP; 1933 A; 437 C; 673 G; 1547 T;
SQ

```

Query Match 4.5%; Score 80; DB 1; Length 4590;
Best Local Similarity 50.9%; Pred. No. 0.0051;
Matches 220; Conservative 0; Mismatches 205; Indels 7; Gaps 1;

```

QY 101 aaaaagtagaacaagaatttattatgaacattagcgcaatccttaagaaaaataaa 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 314 AAAAAAATAATAAACCTAATTATAAAAAAATAAAGTTTCATATATATTTTTTTT 373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 161 agtttattctattttttttaaataacccacacagcgataatcttctgcagtattat 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 374 TTTTCTTATATATATGATATTTTCTGTAGAAAAAATAAAGAAATAA 433
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 221 caagtcggtatttttctattctactaaataattatttgatttgcatttccataa 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 434 TTATTTATATATATAAATATTTTTTTTATATATATATATATATATATATATTA 493
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 281 tctaaattttacatttttttaacaaatttttaaaataactctttaaatttagtatt 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 494 AAAAAAATAAATAAATTTTAAATAATTTTATTTATTTATTTATATATATATATTA 553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 341 tttttatttttagtctaaattataaataattcttgaatttttttgaatttttataatt 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 554 TTTTCTTATATATATATATATATATATATATATATATATATATATATATATAT 613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 401 tagtactaaaaatacaaatatttttccattctcctaagaaaaattcttttttaaaaaa 460
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 614 T-----TAATATATATATATATATATATATATATATATATATATATATATAT 666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 461 attgatttttagtataattttgtgataattgaattgaattgaattgaattgaattga 520
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 667 ATTACTATTTTATATATATATATATATATATATATATATATATATATATATATAT 726
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 521 aaatgaaaaaaa 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 727 AAAAAAATAAATAA 738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 10
Q53480/C
ID Q53480 standard; cDNA; 2503 BP.
AC Q53480;
DT 30-JUN-1994 (first entry)
DE pNX30 xylanase cDNA.
KW xylanase; ruminant animals; fungus; paper; pulp; bagasse;
KW feedstock; rumen; plant fibre; ss.
OS Neocallimastix patriciarum.
FH Key Location/Qualifiers

```

FT cds      1..1935
FT FT        /*tag= a
FT PN        /product= Xylanase.
PD WO9325671-A.
PF 23-DEC-1993.
PF 17-JUN-1993; AU0294.
PR 17-JUN-1992; AU-002985.
PR 29-JUN-1992; AU-003238.
PR 01-APR-1993; AU-008100.
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI Xue GP;
DR WPI; 94-007529/01.
DR P-PSDB; R44529.
DR New recombinant fungal xylanase - used for hydrolysis of xylan in
PT food and pulp and paper industries and for improving ruminant
PT feed efficiency.
PS Claim 3; Figure 3; 45pp; English.
CC The cloned xylanase coding sequence is derived from an anaerobic
CC rumen fungus. The xylanase has high specific activity for the
CC hydrolysis of xylan. It can be used for treating pulps in the pulp
CC and paper industry, for treating bagasse for more efficient disposal
CC or for the treatment of feedstock to improve nutritional value.
CC Genetically modified xylanase genes can also be used for the
CC modification of rumen bacteria to improve plant fibre utilisation by
CC ruminants.
SQ Sequence 2503 BP; 893 A; 389 C; 517 G; 704 T;

```

Query Match 4.4%; Score 77.6; DB 1; Length 2503;
Best Local Similarity 49.4%; Pred. No. 0.011;
Matches 229; Conservative 0; Mismatches 234; Indels 1; Gaps 1;

```

QY 46 taattaaaaattacatttcttcattgcccagaatttttaagaatttagcacattaaaaa 105
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2474 TAATTTTTCATATATCTTTTATTTTAAATTTTAAATTTTAAATTTTCAATTTTAAAT 2415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 106 gtagaacaagaatttataatgtaaacattagcgcaatcccttaagaaaaattaaagt 165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2414 TTTACAAATTTTATATTTTATATAAATTTTAAATTTTTCATATATTTTATTTTACTA 2355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 166 tatctatttttttaacgaaatcccaacagcgataaatcttctgcagttattatcaggt 225
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2354 AATCAATATTTTAAATTTTAAATTTTAAATTTTCAATTTTAAATTTTAAATTTTAAAT 2295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 226 csgtatttttctattctactactaaataattattgaatttgcatttccataatcaa 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2294 CAACACTTTTATATTTTATATTTTAAATTTTCAATTTTAAATTTTAAATTTTAAAT 2236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 aattttacatttttttaacaaatttttaaaataactcttcttaattatagtttttt 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2235 ATTTTAAATTTTACATTTTAAATTTTTCATATATTTTATTTTATATATATTTTAT 2176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 attttttagtcataattataaaattctcttgaaatttttttgaattttttataattagta 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2175 TTTTCTTACAAATTTGTATTTTAAACCTTTTACATTTTAAATTTTTCATATTTCTTCAT 2116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 ctaaaaatacaaaatttttccattctcctaagaaaaattcattttttaaataaataa 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2115 TTTTCTTACAAATTTGTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2056
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 tttttatagataaattgttgataaattgaattgaattgaattgaattgaattgaatt 509
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2055 ATTTTCTACTAAATTTAGTTTTTTAAATTTTAAATTTTAAATTTTAAATTT 2012
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 11
Q28302/C
ID Q28302 standard; DNA; 1511 BP.
AC Q28302;
DT 12-FEB-1993 (first entry)
DE AMEPV tk DNA.
KW Entomopoxvirus; thymidine kinase; non-essential; regulatory sequences;
KW vector; ss.

[illegible]

Db	481	GGAAATGGCCGACACAGCAATACCAAAATTAACCTTGGAAGGATTGTGATACCCAAAAAATC	540
Qy	1477	tttgttaactcgtcaagattataatgataaagcctaaactttatacaagacgcgcatcaa	1536
Db	541	TTTGTAACTCCTCAAGATTATAATGATRAAGCCAAAACCTTTATCAAAAGACGGCGATCAA	600
Qy	1537	aatatgacaatttataaaacctgataaagtttttaggaaaagttgctgttggaagttcttcgg	1596
Db	601	AATATGACAAATTTATAAAACCTGATAAAGTTTGTAGCAAAAGTTGCTGTTGAAGTCTTCTCG	660
Qy	1597	gttttaattgcgaagaataaataaagacatctagatcagaagtcgaaacgaactaaagaaca	1656
Db	661	GTTTTTAATTGCAAGAAAAAATAAAGCATCTAGATCAGAAGTCGAAAAACGAACATAAAGCA	720
Qy	1657	aaactaccaaaatttcatttaatatgatataatcaaacatataaagtcacaagtgtaaaaat	1716
Db	721	AAACTACCAAAATATTTCATTTAAATATGATATCATCAACATATAAAGTCGAAGGTGAANAAT	780
Qy	1717	attaatacaatttttagtaagtcagctgaatgtttacaaagctaagttgtataatcctgat	1776
Db	781	ATTAATAACAATTTTACTAAGTCCACTAATGCTGTACAAAGACTATGTTGATATATCCTGAT	840
Qy	1777	gcc	1779
Db	841	GCC	843

RESULT 4

Q05576
ID Q05576 standard; DNA; 396 BP.
AC Q05576;
DT 19-DEC-1990 (first entry)
DE Sequence encoding swine enzootic pneumonia mycoplasma surface
DE antigen polypeptide.
KW Mycoplasma pneumoniae; enzootic pneumonia; ds.
OS Mycoplasma hyopneumoniae.
PN J02167079-A.
PD 27-JUN-1990.
PF 21-DEC-1988; 322829.
PR 21-DEC-1988; JP-322829.
PA (NISE-) NIPPON SEIFUN KK.
DR WPI; 90-241949/32.
DR P-PSDB; R06277.
PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
PT for early detection and treatment of mycoplasma or enzootic
PT pneumonia of pigs
PS Caltm 2; Page 570; 28pp; Japanese.
CC Sequence encoding polypeptides may be used as probes for the early
CC diagnosis of pneumonia in pigs. The sequences may be used to produce
CC an expression vector and transform a suitable host.
SQ Sequence 396 BP; 145 A; 62 C; 70 G; 119 T.

Query Match	22.0%	Score 392.8;	DB 1;	Length 396;
Best Local Similarity	99.5%	Pred. No. 1.9e-41;		
Matches 394: Conservative	0: Mismatches	2: Indels	0	

Qy	997	gataatgaaaaagttggtggaattacaaggctctttcaacttgcgtcggtgctattaggaaaa	1056
Db	1	GATAATGAAAAAGTTGGCGAAATACAAAGTCTTTTTCAC'TTGGCGGGGTCTATTAGGAAAA	60
Qy	1057	gaagatgggtcttttgattcaattgaatcaaatgaatgaatatcaaatcacatatgcc	1116
Db	61	GAAGATGGTGTCTTTTGATTCAATTGATCAAA'TGAATGAATATCTAANAATCACATATGCCC	120
Qy	1117	caagagacaattctcttttatacaatcgcgggtctcccaagatgataataatcccaatat	1176
Db	121	CAAGAGACAATTCTTTTATACAATCGCGGGTCCCAAGATGATATATATTTCCCAATAT	180
Qy	1177	ttttataatgggtgcaatgaagatgaattaaatgaataattcgcataataaaaa	1236
Db	181	TTTTTAATGGTGCATGAAGTACTTAAAGAA'TTAATGAANAATTCGCAAAATAAAAA	240

[illegible]

RESULT

Q05377
ID Q05577 standard; DNA; 360 BP.
AC Q05577;
DT 19-DEC-1990 (first entry)
DE Sequence encoding swine enzootic pneumonia mycoplasma surface
KW antigen polypeptide.
DE Mycoplasma pneumoniae; enzootic pneumonia; ds.
OS Mycoplasma hyopneumoniae.
PN J02167079-A.
PD 27-JUN-1990.
FF 21-DEC-1986; 322829.
PR 21-DEC-1988; JP-322829.
FA (NISE-) NIPPON SEIFUN KK.
DR WPI; 90-241949/32.
DR P-PSDB; R06278.
PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
PT for early detection and treatment of mycoplasma or enzootic
PT pneumonia of pigs
FS Claim 1; Page 570; 28pp; Japanese.
CC Sequence encoding polypeptides may be used as probes for the early
CC diagnosis of pneumonia in pigs. The sequences may be used to prod
CC an expression vector and transform a suitable host.
CC Sequence 360 BP; 130 A; 52 C; 64 G; 114 T;
SQ

Query Match	20.0%	Score 356.8;	DB 1;	Length 360;
Best Local Similarity	99.4%	Pred. No. 5.2e-37;		
Matches 358: Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	979	gattggctatgttcttcttgataatgaaaagtgggtgaattaccagaagtccttctcacttgcct	1038
Db	1	GATTGGTATGTTCTTTGATAATGAAAAAGTTGGCAATTACAAGGTCCTTTCACCTTGGC	60
Qy	1039	gcgggtctataggaagaaagaatgggtgcttttgattccaaattgatcaaatgaatgaatat	1098
Db	61	CGGGCTCTATTAGGAAGAAGATGGTGGTTTGTGATTCAAATGATCAATGAATGAATAT	120
Qy	1099	ctaaaatcacatatgccccaaagacaaattctcttttatcaaatcgcggggtccccaagat	1158
Db	121	CTAAATTCACATATGCCCAAGAGACAAATTTCTTTTATACAATCGGGGTCCCAAGAT	180
Qy	1159	gataaataattcccaatattttataatgggtgcaatgaagtacttaagaagaataatgaaa	1218
Db	181	GATAAATAATTCCCCAATATTTTTATATGGTGCATGAAGAAGTACTTTAAGCAATTAATGAA	240
Qy	1219	aattcgcaaaataaaaataattgatttatctctctgaagcgaaaaatgctgtttatgtccca	1278
Db	241	AATTTCGCAAAATAAATAATTGATTATTTCTCTGAAGCGAAAATGCTGTTTATGTCCCA	300
Qy	1279	ggatgaataattatggaactgcgggtcaagaagatcccaattctttcttaacaattaacaaagat	1338
Db	301	GGATGAAATTTATGGAACCTCGGCTCAAGCAATCCAAATCTTTCTTCACAATTAACAAAGAT	360

RESULT

RESULTS
T72882/c
ID T72882 standard; cDNA; 19124 BP.
AC T72882;
DT 12-SEP-1997 (first entry)

```

QY 583 gcatcaattattgcatcttggtagcaggtgttgagacagacagaatcaggttcaattct 642
Db 61 GCATCAATTATGCTATTTGTTGAGCAGGTTTGGACAGACAGAAATCAGGTTCGACTTCA 120
QY 643 gattctaaacacacagccgagcgttaaaacataaagtaagtaagtattcttctgaata 702
Db 121 GATTCTAAACCAACAGCCGAGACTCTAAACATAAAGTAAGTAATGATTCTATTCGAATA 180
QY 703 gactaacccgcatcgggataatcctcgatgaattagtgcccaaaaagatatattcttat 762
Db 181 GCATACCGATCCGGATATCTCGATGATTAAGTTCGCCCAAAAAGATATATTCTTTAT 240
QY 763 gttgatgaacacagaggcagcaattcaacaattacaaaaaccaggatgcacaaaaaac 822
Db 241 GTTGATGAACACAGAGCGCAACTCAACAATTAACAAAAACCAGGATGCACAAAAATAC 300
QY 823 tgactcaactcagcaagctaattaaagccagccagcccaaaaggattatttgcacctgaa 882
Db 301 TGACTCACTCAGCAAGCTAATTTAAGTCCAGCCGCAAAAGGATTTATTATGCCCCGAA 360
QY 883 aatggagtgaggttggaactgctgtaatacaaatgctgataaaggaaatccgattgtt 942
Db 361 AATGGAGTGGAGTGGAACTGCTGTTAATACAAATGCTGATAAAGAAATTCGGATTGTT 420
QY 943 gcctatgacgactaattactggtatctgataaaatgatggttatttttttgataat 1002
Db 421 GCCTATGATCGACTAATTACTGATCTGATAAATATGATTGGTATGTTCTTTTGATAAT 480
QY 1003 gaaaaagtgtgtaatacaaggtcttctcaactgctgcgggtctattaggaagaagaat 1062
Db 481 GAAAAAGTGGCGGAATTAACAAGCTCTTCTACTTGGCGCGGGTCTATTAGAAAAAGAAAT 540
QY 1063 ggtgcttttgattcaattgatcaaatgaatgaatctataatcacaatcacaatgcaccaagag 1122
Db 541 GGTGCTTTTGATTCATTTGATCAATGAATGAATGAATATCTAAATCACAATATGCCCCAAGAG 600
QY 1123 acaattcttttatacaatcgcggtgtcccaagatgataataatcccaattttttat 1182
Db 601 ACAATTTCTTTTATACAATCGCGGGTCCCAAGATGATAAATAATTCCTCAATATTTTAT 660
QY 1183 aatggtgcaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1242
Db 661 AATGGTCAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 720
QY 1243 ttatctcctgaagcgcaaatgctgtttatgtccagagatgaatgaatgaatgaatgaatgaat 1302
Db 721 TTATCTCTGAAGCGCAAAATGCTGTTTATGTTCCAGGATGAATTTATGGAATGCGCGGT 780
QY 1303 caaagaatcccaatcttttcttaacaattaaacaagatccagcgaggtggttaataaaatcaaa 1362
Db 781 CAAAGAATCCAATCTTTTCTAACAATTAACAAGATCCAGAGTGGTAAATAAATCAAA 840
QY 1363 gctgttggttcaaaacacagctctatttcaagagattcttgcccaaatgatggaatg 1422
Db 841 GCTGTGGTTCACCAACAGCTCTATTATTTCAAGAGATTTCTTGCCCAATGATGGAATG 900
QY 1423 gccgaacaagcaatcaccataataaactggaaggtttgataccacaaaatctttgta 1482
Db 901 GCGGAACAGCAATACCAAAATTAACCTTGAGGATTTGATACCCCAAAAATCTTTGTA 960
QY 1483 actcgtcaagattataatgaagccaaaacttttatcaagagcgagcgtcaaaaatg 1542
Db 961 ACTGGTCAAGATTATAATGATAAAGCCAAAATCTTTTATCAAGAGCGGATCAAAAATATG 1020
QY 1543 acaattataaacctgataaaagtttttagaagaaggtgctgttggaagttcttcgggtttta 1602
Db 1021 ACAATTTATAACCTGATAAAGTTTATAGAAAAGTGTGCTGTTGAAGTCTTCGGTTTTTA 1080
QY 1603 attgcaagaaaaataaagcatctagatcagagtcgaaacgaaactaaaaacaaacta 1662
Db 1081 ATTGCAAAAGAAAATAAAGCATCTAGATCAGAAATCGAAACGAACTAAAGCAAAACTA 1140

```

```

QY 1663 ccaaatatttctttaaataatgataatcaaacataataaagtaacaggttaaaatattat 1722
Db 1141 CCAAAATATTTCATTAAATATGATAATCAACATATAAGTCAAGGTAAAAATATTAAAT 1200
QY 1723 acaattttagtagccagtaatttttcaaaaagcgaatgttgataaactcctgatccctaa 1782
Db 1201 ACAATTTTAGTAGTCCAGTAATTGTTACAAAAGCTAATGTTGATAATCTCTGATGCCATA 1260

RESULT 3
Q05578
ID Q05578 standard; DNA; 843 BP.
AC Q05578;
DT 19-DEC-1990 (first entry)
DE Sequence encoding swine enzootic pneumonia mycoplasma surface
DE antigen polypeptide.
KW Mycoplasma pneumoniae; enzootic pneumonia; ds.
OS Mycoplasma hyopneumoniae.
PN J02167079-A.
PD 27-JUN-1990.
PF 21-DEC-1988; 322829.
PR 21-DEC-1988; JP-322829.
PA (NISE-) NIPPON SEIFUN KK.
DR WPI; 90-241949/32.
DR P-PSDB; R06279.
PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
PT for early detection and treatment of mycoplasma or enzootic
PT pneumonia of pigs
PS Claim 2; Page 570; 28pp; Japanese.
CC Sequence encoding polypeptides may be used as probes for the early
CC diagnosis of pneumonia in pigs. The sequences may be used to produce
CC an expression vector and transform a suitable host.
SQ Sequence 843 BP; 324 A; 127 C; 139 G; 253 T;

Query Match 46.9%; Score 835; DB 1; Length 843;
Best Local Similarity 99.4%; Pred. No. 4.7e-96;
Matches 838; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 937 attgttgcttgatcgactaattactgactgataaataatgattggttattttttt 996
Db 1 ATTGTGTGCTTATGATCGACTAATTACTGATGATAAATATGATTGTTTCTTTT 60
QY 997 gataatgaaaagtgtgtaattacaaggttttcaactgctgcgggtctattaggaataa 1056
Db 61 GATATGAAAAGTTGGCGAATTACAAGTCTTTCACCTTGGCGGGTCTATTAGGAAAA 120
QY 1057 gaagtgtgttttttgattcaatgatcaaatgaatgaatgaatgaatgaatgaatgaatgaat 1116
Db 121 GAAGATGCTGCTTTTGAATCAATGATCAATGAATGAATATCTAAATATCATATATGCC 180
QY 1117 caagagacaatttttttttttttttttttttttttttttttttttttttttttttttttt 1176
Db 181 CAAGAGACAAATTTCTTTTATACAAATCGCGGTTTCCCAAGATGATAAATATCCCAATAT 240
QY 1177 tttataatgtgcaatgaagtaacttaagaataataaataatgaataatgaataatgaataatgaata 1236
Db 241 TTTTATAATGTCNATGAAAGTACTTAAAGAAATTAATGAAAAATTCGCAAAATAAAAATA 300
QY 1237 attgatttatctctggaaggcgaataatgctgtttatgtcccgaggtgaaattatggaact 1296
Db 301 ATTGATTTATCTCTGAAGCGCAAAATGCTCTTTATGTGTCGCCAGGATGAATATATGGAAT 360
QY 1297 gccggtcaagaatccaatctttttcttaacaataaacaagaatccagcaggtggtaataaa 1356
Db 361 GCCGGTCAAGAATCCCAATCTTTTCTTAACAATTAACAAGATCCAGCAGGTGGTAATAAA 420
QY 1357 atcaaatgctgtgtgtcaaaaaccagctcttatttttcaagattcttgcctcccaaatgat 1416
Db 421 ATCAAGCTGTGTGTTCAAAACAGCTCTTATTTTCAAGAGATTTCTTGCCCAAAATGAT 480
QY 1417 ggaatggccgaacaagaatcaccacaaattaaaaacttgaaaggtttgtatcaccacaaatc 1476

```



```

/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="cytochrome c oxidase subunit II"
/protein_id="AAC47813.1"
/db_xref="PID:g1166532"
/db_xref="GI:1166532"
/translacion="MSTWANLGQDSASPLMEQLIEFHDLILVIMITVLGYLMFM
LFNNYVNFRLHGOIEMITWTILPAIILFIALPSRLYLDEINEPSVLKSIQH
OMWYSVEYSDFNNEIDSYMIPTNEMLDGFRLLDYNRVLPMSQIRILVTAADVI
HSWTVPALGVKVDGTGRLNQTFFINRPGLYGQCSEICGANHSPMPVIESVPVNY
FKWISSNNS"
3768..3838
/gene="mt:ND6"
/product="trNA-ND6"
/db_xref="FlyBase:FBgn0013685"
3840..3906
/gene="mt:ND6"
/product="trNA-Asp (GUC)"
3907..4068
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="ATPase 8"
/protein_id="AAC47814.1"
/db_xref="PID:g1166533"
/db_xref="GI:1166533"
/translacion="MPQMAPISWLLLIIFISITFIILFCSINYSPNSPKSNELKNI
NLNSMNWKKW"
4062..4736
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"

Query Match      5.4%; Score 95.6; DB 37; Length 19517;
Best Local Similarity 49.1%; Pred. No. 0.014;
Matches 283; Conservative 0; Mismatches 289; Indels 4; Gaps 1;

QY  2  tgaataaaatgcactataccagaggagagcagtgatataataataataataataacatt 61
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17242 TAAATGAAAATAATTTTAAATTTTAAATTAATAATAATAATAATAATAATAATAAT 17301

QY  62  ttcttcattgcgcagaatttttaagaatttagtacattaaagaagtagacaacaaagtta 121
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17302 ARAATCTATTCAATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17361

QY  122  taatgtaaacattagcgcaatcccttaagaagaaataaaagttttctctatttttttaa 181
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17362 ATAAATTTTATTATTACTAATATTAAATTAATAATAATAATAATAATAATAATAAT 17421

QY  182  tcgaataccacaggcataaaactttgtcagttatttatcaagtcggtatttttccatta 241
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17422 TT-----TTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 17477

QY  242  ttctactaaaattatttgaaattgtcatttccataactcaaaaatttaccattttttt 301
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17478 TTATATATATATATATATATAGAAAAATTAATTTTAAATAATTTAAATAATAATTT 17537

QY  302  ataacaatttttaaaaattactctttaaattatagttatttttttttttttagtctaaa 361
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17538 TTTAAAAAATTTCTAAATGTATTATTATTATATAAAAAATTTTATATAATAAATCATGTT 17597

QY  362  tataaaaattcttgaaatttttatttgatttttataatttttagtactaaaaatacaaaa 421
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17598 TTTAAAAAATAACAAAAAATTTTAAATAATAATAATAATAATAATAATAATAATAATTT 17657

QY  422  tttttctcttctcaagaaaaattcatttttaaaaaaaatttgattttttatagataaatt 481
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17658 TATTTTTCATTTTAAAAAATAATTTTAAAAAATAATAATAATAATAATAATAATAAAC 17717

QY  482  tgtttgtataattgaatttaacttgatttgaaagggaacaaaatagaacaaaatgcttagaa 541
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 17718 TATATACTAATTAATAAATAATTAAGATATTTTATATATATAATAAATAATTAATAATATATATTA 17777
QY  542 aaaaattcttgattcatcagctatttatgcaactt 577
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 17778 TATATCTAATAATTTAATAAAAAAATTTTAAAAATTT 17813

Search completed: November 13, 1999, 14:28:59
Job time: 4914 sec

```

AUTHORS Clary,D.O., Goddard,J.M., Martin,S.C., Faure,C.M. and Wolstenholme,D.R.
TITLE Drosophila mitochondrial DNA: a novel gene order
JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)
MEDLINE 83090428
REFERENCE 2 (bases 5269 to 5695)
AUTHORS Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.
TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial trna genes
JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)
MEDLINE 83220794
REFERENCE 3 (bases 404 to 5272)
AUTHORS de Bruijn,M.H.
TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code
JOURNAL Nature 304 (5923), 234-241 (1983)
MEDLINE 83245048
REFERENCE 4 (bases 804 to 1778)
AUTHORS Satta,Y., Ishiwa,H. and Chigusa,S.I.
TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species
JOURNAL Mol. Biol. Evol. 4 (6), 538-550 (1987)
MEDLINE 88174373
REFERENCE 5 (bases 5268 to 13619)
AUTHORS Garesse,R.
TITLE Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations
JOURNAL Genetics 118 (4), 649-663 (1988)
MEDLINE 88212147
REFERENCE 6 (bases 441 to 2967)
AUTHORS Satta,Y. and Takahata,N.
TITLE Evolution of Drosophila mitochondrial DNA and the history of the melanogaster subgroup
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
MEDLINE 91088557
REFERENCE 7 (bases 14215 to 14512)
AUTHORS Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W.
TITLE Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods
JOURNAL Science 258 (5086), 1345-1348 (1992)
MEDLINE 93088057
REFERENCE 8 (bases 14917 to 19517)
AUTHORS Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA
JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE 94285822
REFERENCE 9 (bases 1 to 408; 13319 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons
JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE 96423163
REFERENCE 10 (bases 1 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department, Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES
source
/organism="Drosophila melanogaster"
/mitochondrion
/db_xref="taxon:7227"
/note="derived from new and previously submitted sequences; sequence is a composite containing sequences obtained from different Drosophila melanogaster strains"
1. .65
/gene="mt:ND6"
/product="tRNA-Ile (GAU)"
/db_xref="FlyBase:FBgn0013685"
1. .19517
trna
gene

6"
/gene="mt:ND6"
/note="mitochondrial NADH-ubiquinone oxidoreductase chain
6"
/allele=""
/db_xref="FlyBase:FBgn0013685"
complement(97. .165)
/gene="mt:ND6"
/product="tRNA-Gln (UUG)"
/db_xref="FlyBase:FBgn0013685"
171. .239
/gene="mt:ND6"
/product="tRNA-f-met (CAU)"
/db_xref="FlyBase:FBgn0013685"
240. .1265
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC47811.1"
/db_xref="PID:g1166530"
/db_xref="GI:1166530"
/translation="MFNNSKILFTIMIGTLITVTSNWLCAWGLEINLLSFIPL
LSDNNLMSTASLKYFTQVLASTVLLFSSILLKNNMNEINETSMTIIMISALL
LKSGAAPHFHFWPNMMEGLTWMLMAMTQRIAPMLISYINIKYLLISVILSVII
KAIGLNQTSLRKLMFASINHLGWSLSMISESTWILFFYFSLFSLVTFMFNIF
KLFHNLQTSFVNSKILKFTLFMNFSLGGLPPLFGLPKWLVTQQLTLCNQYMLT
IMMSTLTLFFYLICYSAFMNYFENWIMKMNNSINYNWYIMTFEFSIFGLFLI
SLFYFNF"
1264. .1329
/gene="mt:ND6"
/product="tRNA-Tip (UCA)"
/db_xref="FlyBase:FBgn0013685"
complement(1322. .1383)
/gene="mt:ND6"
/product="tRNA-Cys (GCA)"
/db_xref="FlyBase:FBgn0013685"
complement(1403. .1468)
/gene="mt:ND6"
/product="tRNA-Tyr (GUA)"
/db_xref="FlyBase:FBgn0013685"
1470. .1473
/gene="mt:ND6"
/note="ATAA, putative translation initiation codon for
cytochrome c oxidase subunit I"
/db_xref="FlyBase:FBgn0013685"
1474. .3009
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="cytochrome c oxidase subunit I"
/protein_id="AAC47812.1"
/db_xref="PID:g1166531"
/db_xref="GI:1166531"
/translation="SRQLFSTNHKDIGTLYFPGAWAGVGTSLILIRAEHLGPGA
LIGDQIVNVIVTAHAFIMFFMVPIMGGFGLVPLMLGAPDMPAPRMNMSFWL
LPPALLSVSSMVENAGTGTWYVPLSAGIAGHGSVDLAIFSLHLAGISSILGAV
NFITVINRSTGISLDRMPLFVWSVITALLLSLPVLAGAITMLLTDRLNLSFF
DPAGGDPILYOHLEWFFGHPEVYIILPFGMISHIIISQESKKETFGSLGMIYAML
AIGLGFVWAHMFVTGMDVDTYATFATMIAVPTGIKIFSWLATIHGTQLSYSP
AIIWALGFVFTVGLTGVVLANSVDIILHDTYVVAHFYVLSMGAVFAIMAGFI
HWPLFTGLTLNNKWLKSHFTIMEFIGNVITFEPHFLGLAGMPRYSDYPDATYTNWI
VSTIGSTISLGLIFFEFLIWSLSVRSQVIYPIQINSIEWYQNTPPAEHSYSELPL
LTF"
3012. .3077
/gene="mt:ND6"
/product="tRNA-Leu (UAA)"
/db_xref="FlyBase:FBgn0013685"
3083. .>3767
/gene="mt:ND6"
/note="incomplete stop codon, requires polyadenylation"
/codon_start=1

AUTHORS Kaguni, L. S.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of
Biochemistry, Michigan State University, East Lansing, MI,
48824-1318. USA

FEATURES Location/Qualifiers

[illegible]

```

BASE COUNT      2271 a 131 c 74 g 2125 t
ORIGIN
/note="deoxythymidylate stretch"
/db_xref="FlyBase:FBgn0013687"

```

Query Match	5.48;	Score 95.6;	DB 36;	Length 4601;
Best Local Similarity	49.1%;	Pred. No. 0.024;		
Matches 283;	Conservative 0;	Mismatches 289;	Indels 4;	Gaps 1;
Qy	2	tgaaaaaataccactataccagaggaagagcagtatataaaataataaaataccatt	61	
Db	2326	TAAATGAAATATTTTAAATTTTAAATATAAATAAATTTAAATAATATAATTAAT	2385	
Qy	62	tcttcatttcgcccgaatttttaagaatttagtacatttaaagaagtagaacaagaattat	121	
Db	2386	AAATCTATTCTATTATAATATTTAAATTAATAAATAAATAAATTTAAATAATTAATTA	2445	
Qy	122	taagtgaacaattagcgcaatccctaagaaaaataaaagttttatctatttttttaa	181	
Db	2446	ATAAAATTTATTTACTAATATTTAAATTAATAAATAAATAAATTTATTTTTTTTTTTT	2505	
Qy	182	tcgaataccaaccaggcataactcttgcagatttatcaagtcgggtatttttcatta	241	
Db	2506	TT---TTTTAAATTAATTAATTTATATATATTAATAATTTATATATATTGAAATA	2561	
Qy	242	ttctactaaatattatttgaatttcgattccataatccataattttacatttttt	301	
Db	2562	TTTATAATATATATATATATAGAAAAATTTAAATTTATTTAAATAATTTAAATAAATTT	2621	
Qy	302	atacaatttttaaaaattacctcttaatttatagttattttttatcttttagcttaaat	361	
Db	2622	TTTAAAAATTTCTTAAATGTAATTTTATTTTAAAAAAATTTATATATAAATAAATCATGTT	2681	
Qy	362	tataaaattacttgaattttatttgaatttttataatttttagtactaaaaatacaata	421	
Db	2682	TTTTAAAAAATAACAAAAAATTTTAAATAAATAAATTTTATAATGAAATATATATTTATT	2741	
Qy	422	tttttctctactcaagaaaattcatttttttaaaaaaaattgattttcttagtataaatt	481	
Db	2742	TATTTTTTCATTTTTTAAAAAATAAATTTTAAAAAATAAATTTTATTTTAAAAAATAAAT	2801	
Qy	482	tgttgtataattgaattcaacttgatttgaaggggaacaaataaaaaaaattgcttagaa	541	
Db	2802	TATATCTAATTAATTAATTAATAGATAATTTATATATATATAAATTTAAATATATATTA	2861	
Qy	542	aaaaattcttgtattcatcagctatttatgcaactt	577	
Db	2862	TATCTTAATTAATTTAAATAAATAAATTTTAAAAATTTTAAAAATTT	2897	

RESULT 15
DMU37541

LOCUS	DMU37541	19517 bp	DNA	circular	INV	30-DEC-1997
DEFINITION	Drosophila melanogaster complete mitochondrial genome.					
ACCESSION	U37541					
NID	g1166529					
VERSION	U37541.1	GI:1166529				
KEYWORDS	A-T-rich region; ATPase; circular; complete genome; cytochrome b; cytochrome c; cytochrome c oxidase; mitochondrial DNA; NADH dehydrogenase; ribosomal RNA; transfer RNA-Ala; transfer RNA-Arg; transfer RNA-Asn; transfer RNA-Asp; transfer RNA-Cys; transfer RNA-Gln; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His; transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer RNA-Ser; transfer RNA-Thr; transfer RNA-Trp; transfer RNA-Tyr; transfer RNA-Val.					
SOURCE	Drosophila melanogaster.					
ORGANISM	Mitochondrion Drosophila melanogaster					
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE	1. (bases 12511 to 12682)					

MEDLINE REFERENCE AUTHORS	9108557 7 (bases 14215 to 14512) Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W.	trna	/product="tRNA-Cys (GCA)" /db_xref="FlyBase:FBgn0013685" complement(1403..1468) /gene="mt:ND6" /product="tRNA-Tyr (GUA)" /db_xref="FlyBase:FBgn0013685" 1470..1473 /gene="mt:ND6" /note="ATAA, putative translation initiation codon for cytochrome c oxidase subunit I" /db_xref="FlyBase:FBgn0013685" <1474..3009 /gene="mt:ND6" /codon_start=1 /db_xref="FlyBase:FBgn0013685" /transl_table=5 /product="cytochrome c oxidase subunit I" /protein_id="AAC47812.1" /db_xref="PID:g1166531" /db_xref="GI:1166531" /translation="SRQLESTNHNKDIGTLYFIFGAWGMVGTSLILRAELGHPGA LIGDQIYIVYTAHAFIMIFPMVPMIGGFGNWLVPMLGAPMDAPFRNMMSFVL LPPALUSLLVSMVENGAGTWTVPPLSAGIAGGASVDLAIFFSLHLAGISILGAV NFITVINRSTGSLDRMPLFVWSVITALLLSLPVLGALITMLTDRNLNTSFF DPAGGDPILYQHLEFWFCHPEVYIILPGFMIISHIIQSQSGKETFGSLGMIYAML AIGLGFVWAHMFVGMVDVTRAYFTSATMIIAVPTGKIFSLATLHGQLSYSP ATLWALGFVLEFTVGGTGVVLANSVDIILHDTVYVAHFHYVLSMGAVPAIMAGFI HWYPLTGLTLNNKWLKSHFIMFICVNLTFPPQHFGLAGMPRYSIDPDAYTTWNI VSTIGTISLLGILFFFIWESLSVQRQVIYIQLNSSIEWYQNTPTPAHNSYSELPL LTN"
TITLE	Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods	misc_feature	
JOURNAL MEDLINE REFERENCE AUTHORS	Science 258 (5086), 1345-1348 (1992) 93088057 8 (bases 14917 to 19517) Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.	CDS	
TITLE	Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA		
JOURNAL MEDLINE REFERENCE AUTHORS	Mol. Biol. Evol. 11 (3), 523-538 (1994) 9425822 9 (bases 1 to 408; 13319 to 19517) Lewis,D.L., Farr,C.L. and Kaguni,L.S.		
TITLE	Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons		
JOURNAL MEDLINE REFERENCE AUTHORS	Insect Mol. Biol. 4 (4), 263-278 (1995) 96423163 10 (bases 1 to 19517) Lewis,D.L., Farr,C.L. and Kaguni,L.S.	trna	
TITLE	Direct Submission		
JOURNAL MEDLINE REFERENCE AUTHORS	Submitted (03-Oct-1995) Laurie S. Kaguni, Biochemistry Department, Michigan State University, East Lansing, MI 48824-1319, USA	CDS	
FEATURES	Location/Qualifiers 1..19517 /organism="Drosophila melanogaster" /mitochondrion /db_xref="taxon:7227" /note="derived from new and previously submitted sequences; sequence is a composite containing sequences obtained from different Drosophila melanogaster strains" 1..65 /gene="mt:ND6" /product="tRNA-Ile (GAU)" /db_xref="FlyBase:FBgn0013685" 1..19517 /gene="mt:ND6" /note="mitochondrial NADH-ubiquinone oxidoreductase chain 6"		
trna			
gene			
trna			
trna			
CDS			
trna			
trna			
CDS			
trna			
trna			

```

/db_xref="FlyBase:FBgn0013687"
/rpl_type=tandem
2044..2388
/gene="mt:ori"
/note="repeat I-C"
/db_xref="FlyBase:FBgn0013687"
/rpl_type=tandem
2491..2511
/gene="mt:ori"
/note="deoxythymidylate stretch"
/db_xref="FlyBase:FBgn0013687"
2512..2648
repeat_unit
/partial
/gene="mt:ori"
/db_xref="FlyBase:FBgn0013687"
/rpl_type=tandem
2649..3112
/gene="mt:ori"
/note="repeat II-A"
/db_xref="FlyBase:FBgn0013687"
/rpl_type=tandem
3113..3576
/gene="mt:ori"
/note="repeat II-B"
/db_xref="FlyBase:FBgn0013687"
/rpl_type=tandem
3577..4040
/gene="mt:ori"
/note="repeat II-B2"
/db_xref="FlyBase:FBgn0013687"
/rpl_type=tandem
4041..4504
/gene="mt:ori"
/note="repeat II-C"
/db_xref="FlyBase:FBgn0013687"
/rpl_type=tandem
4505..4565
complement(4565..4585)
/misc_feature
BASE COUNT 2271 a 131 c 74 g 2125 t
ORIGIN
Query Match 5.5% Score 97.4; DB 36; Length 4601;
Best Local Similarity 50.8%; Pred. No. 0.016; Matches 271; Indels 4; Gaps 2
Matches 284; Conservative 0; Mismatches 271; Indels 4; Gaps 2
OY 41 taaataaattcaattcatttcattcatttcgacgaattttaagaattagtcacatt 100
Db 2617 TTATATTAAATATTAAATATTAAATATTAAATTTCTATATATATATATATATTAAATATT 2558
OY 101 aaaaagtagaacaagaattcaattcaacattagcgaaccccttaagaaaaattaa 160
Db 2557 CAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2498
OY 161 agtttattcatttttttaacgaaatccaaaccaggcaataacttgcgattat 220
Db 2497 AAAAAAAT-AAATTTTATTTATTTATTTAAATTTAAATTTAAATTTAAATTTTATTTATTA 2439
OY 221 caagcggatttttcattcatttcacctcaaaatatattgaattgcatttccataa 280
Db 2438 TTAGTTATTTAAATTTATTTATTTATTTATTTAAATTTAAATTTAAATTTAAATTTTATTTAA 2379
OY 281 tctaaatttcacattttttataacaacattttaaanaattactcttaatttagtatt 340
Db 2378 TATTATATTATTTAAATTTATTTATTTATTTAAATTTAAATTTAAATTTAAATTTTAAATTTATA 2319
OY 341 ttttatttttagctcaataataaataattacattgaatttgaatttttaaat 400
Db 2318 TATATATATATATATATATATTTTATTTATTTATTTAAATTTT---ATTATTTAAAT 2262
OY 401 tagtactaaaaatacaaatatttttctattctaagaanaattcaatttttaaaaaa 460

```

Db	2261	AATTATATATAAATAAGTTTATTAAGATATATTTAAATAAAATTTTTTTTTAAAAAA	2202
Oy	461	attgattttttagataattgttgatataactgaacttaactgattggaaaggaca	520
Db	2201	AATATTTTTTAAAGTTTAAATTAATATAATTAATTAATGAATGCGGAATTAATTTATT	2142
Oy	521	aattcaaaaaaatctgtgaaaaaacctcgttatcttcgcacatgatcgcaactcgc	580
Db	2141	TCAATTTCACATATATATATATATATATATATATACAAATTAATTAATTCAGATTAGCAT	2082
Oy	581	tgcacataactatgccatt	599
Db	2081	TAAATATAATTTATTTTATT	2063
RESULT	10		
LOCUS	DMU37541/c		
DEFINITION	DMU37541	19517 bp	DNA circular INV 30-DEC-1997
ACCESSION	U37541		
VERSION	g1166529		
KEYWORDS	GI:1166529		
	AT-rich region; ATPase; circular; complete genome; cytochrome b;		
	cytochrome c; cytochrome c oxidase; mitochondrial DNA; NADH		
	dehydrogenase; ribosomal RNA; transfer RNA-Ala; transfer RNA-Arg;		
	transfer RNA-Asn; transfer RNA-ASP; transfer RNA-Cys; transfer		
	RNA-Gln; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His;		
	transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer		
	RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer RNA-Ser;		
	transfer RNA-Thr; transfer RNA-Trp; transfer RNA-Tyr; transfer		
	RNA-Val.		
SOURCE	Drosophila melanogaster.		
ORGANISM	Mitochondrion Drosophila melanogaster		
	Eukaryote; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Braconytera;		
	Muscomorpha; Ephydroidea; Tephrochilidae; Drosophila.		
REFERENCE	1 (bases 12511 to 12682)		
AUTHORS	Claury,D.O., Goddard,J.M., Matlin,S.C., Fauron,C.M. and		
	Wolstenholme,D.R.		
TITLE	Drosophila mitochondrial DNA: a novel gene order		
JOURNAL	Nucleic Acids Res.	10 (21), 6619-6637	(1982)
MEDLINE	83090428		
REFERENCE	2 (bases 5269 to 5695)		
AUTHORS	Claury,D.O., Wahleithner,J.A. and Wolstenholme,D.R.		
	Transfer RNA genes in Drosophila mitochondrial DNA: related 5'		
	flanking sequences and comparisons to mammalian mitochondrial tRNA		
	genes		
JOURNAL	Nucleic Acids Res.	11 (8), 2411-2425	(1983)
MEDLINE	83220794		
REFERENCE	3 (bases 404 to 5272)		
AUTHORS	de Bruijn,M.H.		
TITLE	Drosophila melanogaster mitochondrial DNA, a novel organization and		
	genetic code		
JOURNAL	Nature	304 (5923), 234-241	(1983)
MEDLINE	83245048		
REFERENCE	4 (bases 804 to 1778)		
AUTHORS	Saito,Y., Ishiya,H. and Chigusa,S.I.		
TITLE	Analysis of nucleotide substitutions of mitochondrial DNAs in		
	Drosophila melanogaster and its sibling species		
JOURNAL	Mol. Biol. Evol.	4 (6), 638-650	(1987)
MEDLINE	88174373		
REFERENCE	5 (bases 5268 to 13619)		
AUTHORS	Garesse,R.		
TITLE	Drosophila melanogaster mitochondrial DNA: gene organization and		
	evolutionary considerations		
JOURNAL	Genetics	118 (4), 649-663	(1988)
MEDLINE	88212147		
REFERENCE	6 (bases 441 to 2967)		
AUTHORS	Saito,Y. and Takahata,N.		
TITLE	Evolution of Drosophila mitochondrial DNA and the history of the		
	melanogaster subgroup		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	87 (24), 9558-9562	(1990)

```
FEATURES
source
    location/Qualifiers
    1..4601
    /organism="Drosophila melanogaster"
    /mitchondrion
    /strain="Oregon-R"
    /db_xref="taxon:7227"
    /dev_stage="embryo"
    1..4601
gene
    /gene="mt:ori"
    /note="mitochondrial origin"
    allele=""
    /db_xref="FlyBase:FBgn0013687"
    650..1022
repeat_unit
    /gene="mt:ori"
    /note="Repeat I-A"
    /db_xref="FlyBase:FBgn0013687"
    /rpt_type=tandem
    1023..1360
repeat_unit
    /gene="mt:ori"
    /note="Repeat I-B1"
    /db_xref="FlyBase:FBgn0013687"
    /rpt_type=tandem
    1361..1705
repeat_unit
    /gene="mt:ori"
    /note="Repeat I-C/A"
    /db_xref="FlyBase:FBgn0013687"
    /rpt_type=tandem
    1706..2043
    /gene="mt:ori"
    /note="Repeat I-B2"
```



```

CDS
1030..2484
/gene="MAL3p2.1"
/note="predicted using hexExon: MAL3p2.1 (PFC0160w),
Putative Ankyrin repeat protein, len: 285 aa"
/codon_start=1
/protein_id="CAB38988.1"
/db_xref="PID:e1407809"
/db_xref="PID:g4493879"
/db_xref="GI:4493879"
/translation="MDNESLSEVLEKQVEIITKAKYQDFIRLQILQIPLLNNDIE
MLASINILHWACVGTETLVKLIISFCIDIEKEDLYNDPTIYAIKNSNVEIYLLLI
KFCGISILFKHRRRSPFLTAICENEDKILHALHLLIYMGVSLSEQNEHGQTA
LFUSVKNNTISLQWLTKEVNLNRHDFYGVTVLHIAVRHCDIDILRLDCYGLNMV
YSSIEKNITFQOLCIKRYFLVYILLKWLQNKICSLKICKIYKIYFWFFAIL
NLIVYENIAHSFIIINKYHFSLVITWFFQOFLWCMYLFKSPGYKEHNFKNKK
NIOSSYKNSPNMGYCTKNAEQVLNTEREIQNKLLISTNFWLTPINHDOALS
NKYNDILINLEYKLSYQVSOERINSLDEYRNALLYNQNPVRKSNKYLIVSTVL
HIYIYIYIFPEFFFEYICDLQYKTSKSSLC"
5344..13329
/gene="MAL3p2.2"
join(5344..12004,12737..13329)
/gene="MAL3p2.2"
/note="predicted using hexExon: MAL3p2.2 (PFC0165w),
Hypothetical protein, len: 1676 aa"
/codon_start=1
/protein_id="CAB38989.1"
/db_xref="PID:e1407810"
/db_xref="PID:g4493880"
/db_xref="GI:4493880"
/translation="MREKLIKLEKENKNRNNIKVRFDCHEMNTYSCQVNVKN
KYENDFWKEYTGNDITIKRSQVEEKKWIEDNPKRESLEFYPYRESNRY
FLINSEINLNGSCVYNNDLIYIIHNNRYEHMNPRTKKKKKKRKTRETRKK
RTLHNNKDTFGSVTSMNDEKYMKYQDYHKDDDDYDDEDDDDYDDEDDDDY
EDDDDDYDDEDDYNNYSNYSNYSNYSNYSNYSNYSNYSNYSNYSNYSNYSN
RKKKEKNSIDKRNLIYSDDDSDNNYEHFTHINSDFLSNDHIEKENNLIQN
EQFINDDVIFRKEFKHLLFVWSDSCNSHWIVKKSILNELYITRSQKNDINDVL
IDVYIMHRIYNTCTRIYINDDFLLNNKTKELGVQNVASHINNTNNMNNNDIH
NVNNIYNNRHLKRIEKKKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
KYKLLIKKMTLHLFNGQTDINILFSHNACLSNKYLIDENYSLLNNYNNLKS
EHLINTSRKKNYTLQDNTNTYNSNTNITKNTNNNNNNNNNNNNNNNNNNNN
LTKIYENYSTYLSYTKECULLIDILYILNVTGNLIYIKVYKDYENYVNLISNI
VOKKNEKRAKSYKNIYDHSKENNRKQNSIYISFSQKNYKNTNMNNMLHAGEE
IPDQSSNIISSNWSNNSKSHNSATSDIPMRNNNNNNNNNNNNNNNNNNNN
FPYKLIHLPLNVNENNSIEAREILYIGYIKTKQFLEINRKSQSCNFTFDI
FCCCLNSIAHLINRIEENVRNINFTKIMKNSTNNNNNNNNNNNNNNNNNN
NDDLSCLFDEYKLTCTDPEKKIFFSNLPKNIQDFNLFNINIEKAEDISLYK
IKTSLPLQIGKLANKIINKI1KKKINNSKYLIDLINLOEYLPDIDINKVLVYL
YIDLSTNWGAKIILNDKVPKFLKLSNHFITGKIIDLVSACSKILNIQKRTSYL
IDQNEYKNINENIDINDMTNQNNRSFLPKDVIYKRIYNNINNDQYNSDHRNPY
FYNDLYKSERQTOGMNKQIINGKRYNYMNPYEDYEYEHNNKDEKSCFLIYDND
KRYVENLIHNOHLIYASKMPELYIKKIDIKELKHFFFFFLQISDLKLYFYLSEY
LEKLYHYKKNILKLNKFPDISIRSSVLYHLAKDYNAVATSDIFSMDVNLJIDL
RKQISNEQOGLNKGANNYEYVHDGDDDDYNNNGNEYNNYEEGYNDEERNYL
KNNDYDQEQADDEEDDDDDYDNDDEYKEYEYENDGNDNYDEADSDHNNNTN
ENYHDQVNNENKNNPLHNDNTNQKKHNEYKYTKNERIIRIKNAELKNKGAJ
KKKKTKKKKYSKLNELRNKNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NFILKTNHNNVPSNNNVEYKGLIISKYKMPYFNFINITFKYTLFVYLNCKY
IEHKLSPVLMHVNENINNECTNNLMICHTRECMLHFKCYLHYHQNDVIKSEY
NNNNKLEWLELIEDDIIHINTYLNILKYSFIMDNNTIICILKLSISHIFRHLK
FNFQNEQIENITHNGVNSVNLNPNTHINNHPHQNNNNNNNNNNNNNNNNNN
LDRKAYISQNTQIKYDKHKNFFLLLEYVNNNDDEVAHFLIKDLNFTYNNKYK
LSYQNTSFNDNINSEYVDFNNDNDSDNTQVTKENVGTYNNNNYKYNHMTYS
NNLEGSVOIKRKKLLMLNDNRNNIDLNHTNIDNNNNIINLMKNNGYNNIYRENI
TSEIKNHSPEYRNIIITNHTINNGNSVHQINYSLSHRESNNNNNNNNNNNN
NLNDKSNFKKNYNNINDNNNNYHTNNNSPKSNKEKNNNNNNNNNNNNNNNN
ILANNKDTILSENPYLUKRNHIFQNKNDHTPSPLHNNNNNNNNNNNNNNNN
FNTSYNNNNKNNIDNIDNSRSYIYIYIYIYIYIYIYIYIYIYIYIYIYIYIY
TYPNDKTAEYKANTMYLYYYQNNELAKIDSFKYGLFFGASQCVTDID
NLSEKKNFNFVEIYIPDFTQYEQYSLKNTNFVALPDNYLYICKVQKLNLPF
MLTTPNNNILLKDAALIKTDEYINNLKSLIKNYI1HPKTFQFNRRFFDLFRN"
complement(13855..15201)
/gene="MAL3p2.3"
/note="predicted using hexExon: MAL3p2.3 (PFC0170c),
Lipoamide acyltransferase, len: 449 aa: Similarity to
lipoamide acyltransferases. C.elegans lipoamide
acyltransferase (WP:2K669.4) BLAST Score: 382, sum P(1) =
9.2e-57; 33% identity in 430 aa overlap."
/codon_start=1
/protein_id="CAB38991.1"
/db_xref="PID:e1407812"
/db_xref="PID:g4493882"
/db_xref="GI:4493882"
/translation="MFVKNVLNLRIBEGSKFGHYLNTSAIHFKIVKCKLFDIGEG
YFCEITDHHKNEGDQVSEMSLTVQSDKAADITISNGVLYVKKYINENDMLKVG
ISEYDITDDEDEEVEEKEENKKEDESDLSNDYISNDYIKASPCVKRRAK
EYKVNLVKGDFYFNKNTSLDELELYNNVKNYSNNINNNNDMDIIEEVLGKIKLA
MCKSNESLOVPLEHLMKCLINLLIKMKSEYKQOQNLQKENTITCILLKLSN
VLKEPILNSFNKNTYMYKNHNISIAVDTPHGLVLPNNKNNVONKNDIQLDILK
SLDRKANNQDLKSDITNGTITVSNGAISGTFATPIVDFNOACILIGIKMEKKLL
DESNNLSNLDIVADTINFTFGADHYIDGATLAQFSKMLKMNINENASLGPLE"
17040..18821
/gene="MAL3p2.4"
17040..18821
/gene="MAL3p2.4"
/note="predicted using hexExon: MAL3p2.4 (PFC0175w),
Hypothetical protein, len: 594 aa: Similarity with a
number of hypothetical proteins from model organisms.
C.elegans hypothetical protein (TR:O01963) BLAST Score:
361, sum P(1) = 4.0e-47, 35% identity in 227 aa overlap."
/codon_start=1
/protein_id="CAB38990.1"
/db_xref="PID:e1407811"
/db_xref="PID:g4493881"
/db_xref="GI:4493881"
/translation="MNIPEQNKNDQINVTIEYKDKMKVVKIPINDTKNNIINN
YDINNDKYNISDCYLMNNEGDLNLYIKKFNQENNLNLYKLSHLEKME
LAILEINLIQDYAYHFKAYFNSNDELLENFNNSKVLTKLTKLSHLEKEDILN
NNNNINDPEIKKQKEQVHNCQKEKIFLFECDIEDTYTLSCGHKYSKECLKN
YKTLSHNDDEDIITKCIDLCKNICIKNDKNIKEQDYKLYVLLHYIKKSK
DLKCPNKPICYIIOSVLNNNNVTCGVCYFCPELFEHRLPLCSYIKKWELENN
DDHKNHTHATKCPNCKNPLEKSGCMVYKICGYSFCWCLDNWKNHKGGFYKCN
KYLENYSKINQKOKKDKKDDVTYDDEKEDTDKNDNDNNNNNNNNNNNNNN
ILNRYHFKSFNDSHQYAEFNSHTQLLFLNFCNKNYINHLKMKFFEDAIQIILCR
KILRWYSYAFYFNNKSDQKHLFEYHOGLEKLNLDILQTKTEDINLQFNNTDNDT
VDDIOQITOMIDIFFKTCFEMNNFV"
complement(21152..23273,23390..23604,23713..23862,
23997..24194)
/gene="MAL3p2.5"
/note="predicted using hexExon: MAL3p2.5 (PFC0180c),
Hypothetical protein, len: 895 aa"
/codon_start=1
/protein_id="CAB38992.1"
/db_xref="PID:e1407813"
/db_xref="PID:g4493883"
/db_xref="GI:4493883"
/translation="MFNACOLKSNCCNEAEEMROSMNGKVEDLYGEBHIINENSDLKK
TIEGPDLDKAVEIGQHTEREVAITAQPDIVTRTVVEFPVPTIETVPKIVYEK
IEKVPRTAKYKVEKVEPVKVDKIVDPDQICLKIYKVEIQENILKRPVPHKK
FVEKTVPEVPKEMKRFHEVEYKIRYIPDKASKTKKNETTACINENDETNNE
EGRVBERGYTEHPNLNRLNGARLIMNGPVVQOEGIIIEHTTTEHTSITSSGTREN
VSCFCNNQVGNQTCYSLSGSCNCPRTGELKTEKTNLHNPFRQVFKPKVKN
TEVQKHVPLSDVPVPVYKPVVYVVEVPLKFDTPVPVPRKIIIPKIKWISDVY
QVDCIKKAPYLKIQDVIRPDCVDIKYKWEKACAVNPNELAQDDVHAMVRNAH
LAERKKYGEFYPIQYKSNHIYICMTYICALLYIFNYFFYFFFFFILLDEEENK
MEENSTNEECKEEIIYLEDEERFINKNENINNTKEIKSDNGKINENKKEIKNL
NENINKNKEINNSKGEIYNNNSKGEIYNNNSKGEIYNNNSKGEIYNNNSKGEI
SVTKNNHNIYNNIINDFAHNDKEKVAIEEHFVRKNIEKAEKADIFNYSKEL
REFKDNQNLQNYNDNLNENKYEIYFENSDFENMEKEDIEKNKNTSKYILEEQ
LTASIPSHPLAMTILQNKWIKQTDLTKHLEIYNDIFIRASVANNYNSLGSNIYSDIM
RNNEFLKSNAPLISPFSPGNRNLENYNNLIQNLDQONKSNYIEHMHNNNOINDYD
NNYENFENHQIKLYEQIIEQNEKSKNCCNFCDCR"

```

gene

FT mat_peptide 1. .360
FT /product='surface antigen of swine enzootic

SOURCE Mycoplasma hyopneumoniae.
 ORGANISM Mycoplasma hyopneumoniae
 Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 Mycoplasmataceae; Mycoplasma.
 REFERENCE 1 (bases 1 to 843)
 AUTHORS Nunofuji,S., Okumura,H., Mise,S., Seto,Y., Kanda,K. and Sekiguchi,S.
 TITLE SURFACE ANTIGEN DNA OF SWINE EPIDEMIC PNEUMONIA MYCOPLASMA, SURFACE ANTIGEN POLYPEPTIDE AND DIAGNOSIS USING SAME DNA AND POLYPEPTIDE
 JOURNAL Patent: JP 1990167079-A 3 27-JUN-1990;
 NIPPON FLOUR MILLS CO LTD
 COMMENT OS Mycoplasma hyopneumoniae
 PN JP 1990167079-A/3
 PD 27-JUN-1990
 PF 21-DEC-1988 JP 1988322829
 PI NUNOBUJI SATOSHI, OKUMURA HAJIME, MISE SHIZUO, SETO YASUHIRO, KANDA KUMIKO, SEKIGUCHI SATORU
 PC C12N15/11,C07K13/00,C12P21/02,G01N33/569//A61K39/00,
 (C12P21/02, PC C12R1:19);
 CC topology: Linear;
 CC strandedness: Single;
 CC *source: library= lambda gt11 library;
 CC *source: clone=pKUM3;
 FH key Location/Qualifiers
 mat_peptide 1..843
 /product='surface antigen of swine enzootic pneumoniae'
 mycoplasma'
 /gene='M.hp46kd antigen gene'.
 Location/Qualifiers
 1..843
 /organism='Mycoplasma hyopneumoniae'
 /db_xref='taxon:2099'
 BASE COUNT 324 a 125 c 141 g 253 t
 ORIGIN

Query Match 46.9%; Score 835; DB 5; Length 843;
 Best Local Similarity 99.4%; Pred. No. 4.3e-76;
 Matches 838; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 937 attttgcctatgacataattactggtatctgataataatgattggttcttttt 996
 |||||
 Db 1 ATTGTGCTATGATGACATAATTACTGGATCTGATAAATGATGATGTTCTTTT 60
 |||||
 Qy 997 gataataaaagtgtgtaattcaaggctttcacttctgctggtctattaggaaa 1056
 |||||
 Db 61 GATAAGAAAAGTGGCAATTACAAGGCTTTTCACTTGGCGGGTCTATTAGGAAA 120
 |||||
 Qy 1057 gaagatggtgcttttgattcaaatgatcaaatgaatgaatctcaaatcacatatgcc 1116
 |||||
 Db 121 GAAGATGCTGCTTTGATTCAATGATCAATGAATGAATCAATCAATCAATGATGCC 180
 |||||
 Qy 1117 caagagacaattctttttatacaatcgcggttcccaagatgataataatcccaatat 1176
 |||||
 Db 181 CAAGAGACAATTCTTTTATACAAATCGCGGTTCCCAAGATGATAAATATCCCAATAT 240
 |||||
 Qy 1177 ttltataatggtcaatgaagtactaaagaattaaagaattgaataattgcgaataaata 1236
 |||||
 Db 241 TTTTATATGTCGAATGAAGTACTTAAAGAAATTAATGAATAATCGCAAAATAAATA 300
 |||||
 Qy 1237 attgatttatctctgaaggcgaataatgcttttattgtcccgagatgaattatggact 1296
 |||||
 Db 301 ATTGATTATCTCTGAAGCGCAATGCTGTTTATGTCCAGGATGAATATTGGAACT 360
 |||||
 Qy 1297 gccggtcaagaatcccaatctttttatacaataatacaagaatccagcaggtggttaataa 1356
 |||||
 Db 361 GCCGGTCAAGAATCCCAATCTTTTCTAACCAATTAACAAAGATCCAGCAGGTGGTAATAA 420
 |||||
 Qy 1357 atcaaaagtgtgttcaaaacagcttctattttcaaaagatttctgtcccaaatgat 1416
 |||||
 Db 421 ATCAAGCTGTGTTCAAAACCAAGCTCTTATTATTTCAAGAGATTCTTGCCCCCAATGAT 480
 |||||

Qy 1417 ggaatggccgaacaagcaatcaccataataaacttgaagggtttgatcccaaaaatc 1476
 |||||
 Db 481 GGAATGGCCGAACAAGCAATCACCAAATTAATAACTTTGAAGGATTTGATACCCCAAAATC 540
 |||||
 Qy 1477 ttgttaactcgtcaagattataatgataaagccaaaacttttatcaaaagcggcatcaa 1536
 |||||
 Db 541 TTTGTAACTGGTCAGATTATATGATAAAGCCAAACTTTTATCAAAAGCCGCGATCAA 600
 |||||
 Qy 1537 aaatgacaattataaactgataaagttttaggaagaagtgtgctgtgaagtcttcgg 1596
 |||||
 Db 601 AATATGACAAATTTATAAACCCTGATAAAGTTTAGGAAGATTGCTGTGAAAGTTCTTCGG 660
 |||||
 Qy 1597 gtttaattgcaaaagaaaaataaagcatctagatcagaagtcgaaagcaataaaagca 1656
 |||||
 Db 661 GTTTTAATTCGAAAGAAAAATAAAGCATCTAGATCAGAAAGTCGAAACCACTAAAAGCA 720
 |||||
 Qy 1657 aaactcccaaatattttcatttaataatgataatacaacataataaagatacaaggtaaaaat 1716
 |||||
 Db 721 AACTACCAAAATATTTTCAATTAATATGATAATCAACATATAAAGTGAAGGTAAAAAT 780
 |||||
 Qy 1717 attaatcaatttttagtccagtaattgttacaagaagtaattgttataatctgat 1776
 |||||
 Db 781 ATTAATACAAATTTTAGTAAGTCCAGTAATTGTTACAAAAGCTAAATGTTGATATCTGAT 840
 |||||
 Qy 1777 gcc 1779
 |||||
 Db 841 GCC 843

RESULT 4
 E02474
 LOCUS DNA encoding surface antigen of swine enzootic pneumoniae 29-SEP-1997
 DEFINITION mycoplasma.
 ACCESSION E02474
 NID 92170705
 VERSION E02474.1 GI:2170705
 KEYWORDS JP 1990167079-A/1.
 SOURCE Mycoplasma hyopneumoniae.
 ORGANISM Mycoplasma hyopneumoniae
 Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 Mycoplasmataceae; Mycoplasma.
 REFERENCE 1 (bases 1 to 396)
 AUTHORS Nunofuji,S., Okumura,H., Mise,S., Seto,Y., Kanda,K. and Sekiguchi,S.
 TITLE SURFACE ANTIGEN DNA OF SWINE EPIDEMIC PNEUMONIA MYCOPLASMA, SURFACE ANTIGEN POLYPEPTIDE AND DIAGNOSIS USING SAME DNA AND POLYPEPTIDE
 JOURNAL Patent: JP 1990167079-A 1 27-JUN-1990;
 NIPPON FLOUR MILLS CO LTD
 COMMENT OS Mycoplasma hyopneumoniae
 PN JP 1990167079-A/1
 PD 27-JUN-1990
 PF 21-DEC-1988 JP 1988322829
 PI NUNOBUJI SATOSHI, OKUMURA HAJIME, MISE SHIZUO, SETO YASUHIRO, KANDA KUMIKO, SEKIGUCHI SATORU
 PC C12N15/11,C07K13/00,C12P21/02,G01N33/569//A61K39/00,
 (C12P21/02, PC C12R1:19);
 CC strandedness: Single;
 CC topology: Linear;
 CC *source: library= lambda gt11 library;
 CC *source: clone=pKUM1;
 FH key Location/Qualifiers
 mat_peptide 1..396
 /product='surface antigen of swine enzootic pneumoniae'
 mycoplasma'
 /gene='M.hp46kd antigen gene'.
 Location/Qualifiers
 1..396
 /organism='Mycoplasma hyopneumoniae'
 /db_xref='taxon:2099'

FEATURES
 source

RESULT 2

E04787 1260 bp DNA PAT 29-SEP-1997
 LOCUS gDNA encoding surface antigen of Mycoplasma hyopneumoniae.
 DEFINITION E04787
 ACCESSION 92172983
 NID 92172983
 VERSION E04787.1 GI:2172983
 KEYWORDS JP 1993091882-A/1.
 SOURCE Mycoplasma hyopneumoniae.
 ORGANISM Mycoplasma hyopneumoniae
 Bacteria: Firmicutes; Bacillus/Clostridium group; Mollicutes;
 Mycoplasmataceae; Mycoplasma.
 1 (bases 1 to 1260)
 Seta, Y., Nunofuji, S., Mize, S., Matsuo, K. and Tsuna, M. .
 SURFACE ANTIGEN DNA OF SWINE ENZOOTIC PNEUMONIA MYCOPLASMA, DNA
 FRAGMENT FOR PRIMER, RECOMBINANT PEPTIDE FOR SURFACE ANTIGEN AND
 DIAGNOSIS OF SWINE ENZOOTIC PNEUMONIA USING THE SAME
 Patent: JP 1993091882-A 1 16-APR-1993;
 NIPPON FLOUR MILLS CO LTD
 COMMENT OS Mycoplasma hyopneumoniae
 PN JP 1993091882-A/1
 PD 16-APR-1993
 PF 05-AUG-1991 JP 1991195479
 PR 27-AUG-1990 JP 90P 224945
 PI SETO YASUHIRO, NUNOFUJI SATOSHI, MISE SHIZUO, MATSUO KANAKO,
 TSUNA MIKA
 PC C12N15/31, A61B10/00, C07K13/00, C12N15/10, C12N15/11, C12Q1/04, PC
 C12Q1/68,
 PC G01N33/569/C12P21/02, (C12N15/31, C12R1.35), (C12Q1/68,
 C12R1.35), PC (C12P21/02,
 PC C12R1.19);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 FH Key Location/Qualifiers
 FH CDS 1..1260
 FT /product='surface antigen of Mycoplasma FT
 FT hyopneumoniae'.
 FEATURES
 Location/Qualifiers
 1..1260
 /organism="Mycoplasma hyopneumoniae"
 /db_xref="taxon:2099"
 BASE COUNT 477 a 202 c 215 g 366 t
 ORIGIN
 Query Match 69.9%; Score 1245.6; DB 5; Length 1260;
 Best Local Similarity 99.3%; Pred. No. 3e-117;
 Matches 1251; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 523 atgaaaaaatcgttagaaaaattctgtattcatcagctatttatcaacttcgctt 582
 Db 1 ATGAAAAAATGCTTAGAAAAAATCTTGTATTCATCAGCTATTATTCGAACCTTCGCTT 60
 QY 583 gcatcaattattgatttgcagcaggttggacagacagaaatcaggttcaacttct 642
 Db 61 GCATCAATTATTGCATTGTTGTCAGCAGGTGTTGGACACACAGATCAGGTCGACTTCA 120
 QY 643 gattctaaaccacaaagccgagacgttaaacataaaagtaagtaagtattcttcgaata 702
 Db 121 GATTCTAAACCAACAGCCGAGACTCTAAACATATAAGTAAGTAATGATTCTTATTCGAATA 180
 QY 703 gcaataaccgactcggataatcctcgatgaattagtgcccaaaaagatattattcttat 762
 Db 181 GCACTAACCGATCCGGATATCTCGATGAATTAGTGTGCCAAAGATATTATTCTTAT 240
 QY 763 gttgtgaacacagagcagcaacttcaacatacaaaaaaacagagatgcacaaataac 822
 Db 241 GTTGATGAACAGAGCCACACTTCAACATTACAAAAACAGGATGCCAAATAAC 300
 QY 823 tgactcaactcagaagctaatttaagccagccagcgcgaaggaattattattgcccctgaa 882

Db 301 TGACTCACTCAGCAGCTAAATTAAAGTCAGCGCCAAAAGGATTTATTATTCCCTCGAA 360
 QY 883 aatgaagtgaggttggaactgctgttaatacaattgctgataaaggaattccgattgtt 942
 Db 361 AATGAAGTGGAGTTGGAACTGCTGTTATACAAATGCTGATAAAGGAATCCGATTGTT 420
 QY 943 gcctatgacgactaattactggtatctgataaataatgattggtgattgttcttttataat 1002
 Db 421 GCCTATGATCGACTAATTACTGATCTGATAAATATGATTGTTGTTCTTTTGATAAT 480
 QY 1003 gaaaaagttggtgaattacaaggttcttcaactgtgctggtgcttattaggaaaaagaat 1062
 Db 481 GAAAAAGTTGGCGAATTAACAAGTCTTTTACATGCGCGGGTCTATTAGGAAAAAGAAT 540
 QY 1063 gctgttttgcattgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1122
 Db 541 GGTGTTTTTGATTCAATTGATCAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 600
 QY 1123 acaattctcttttatacaaatcgcggttcccaagatgataataatcccaatttttat 1182
 Db 601 ACAATTCTTTTATACAAATCGCGGGTCCCAAGATGATAATAATCCCAATATTTTAT 560
 QY 1183 aatggtgcaatgaagtaacttaagaataataataataataataataataataataatgatt 1242
 Db 661 AATGTTGCAATGAAAGTACTTTAAAGAAATTAATGAAAAATTCGAAAAATAAATAATTGAT 720
 QY 1243 ttatctctgaagcgaaaaatgctgtttatgctccagatgataatggaactgcccgtt 1302
 Db 721 TTATCTCTGAAGCGGAAATGCTGTTTATGTCAGATGAAATTTATGGAACTGCCCGT 780
 QY 1303 caaagaatcccaatctttctcaaatcaagaatcagcaggtggttaataaaatcaaa 1362
 Db 781 CAAAGAATCCAATCTTTTCTAACAAATTAAACAAGATCCAGCAGGTGGTAATAAATAATCAAA 840
 QY 1363 gctgttggttcaaaaccagctcttattttcaagaattcttgcgcccaaatgatggaatg 1422
 Db 841 GCTGTTGTTCAAAACCAGCTTCTATTTTCAAGGATTTCTTGCCCAAAATGATGGAATG 900
 QY 1423 gccgaacagcaatcaccacaaataaaacttgaaaggttgcatacccaaaaatctttgta 1482
 Db 901 GCCGAACAGCAATCACCAAAATTAACCTTGAAGATTGATACCCCAAAATCTTTGTA 960
 QY 1483 actgtcagaattataatgataaagccaaactttatcaagacgagcgatcaaaatag 1542
 Db 961 ACTGTCAGATTATAATGATAAAGCCAAACTTTTCAAGAGCGGCGATCAAAATATG 1020
 QY 1543 acaattataacacctgataaagtttttagaaaagtgtgctgtggaagttcttcgggtttta 1602
 Db 1021 ACAATTTATAAACCTGTATAAAGTTTAGGAAAAGTTGCTGTGAAGTCTTTCGGGTTTTA 1080
 QY 1603 attgcaaaagaaaaataagcatctagatcagaagtcgaaaaacgaactaaaaagcaaaacta 1662
 Db 1081 ATTGCAAAAGAAAAATAAAGCATCTAGATCAGAAGTCGAAAAACGAACATAAAGCAAAACTA 1140
 QY 1663 ccaaatattctattcaatatgataatacaacataaagaacaggttaaaatattat 1722
 Db 1141 CCAATATTTCATTAAATATGATAATCAACATATAAAGTGCAGAGGTAATAAATAATTAT 1200
 QY 1723 acaatttagtaagtcagtaattgtttacaaaagtaattgttgataatccttgatgcctaa 1782
 Db 1201 ACAATTTTAGTAAGTCCAGTAATTGTTACAAAAGCTAATGTTGATAATCTCTGATGCCTAA 1260
 RESULT 3
 E02476
 LOCUS E02476 843 bp DNA PAT 29-SEP-1997
 DEFINITION DNA encoding surface antigen of swine enzootic pneumoniae
 mycoplasma.
 ACCESSION E02476
 NID 92170707
 VERSION E02476.1 GI:2170707
 KEYWORDS JP 1990167079-A/3.

JOURNAL for arginine
MEDLINE J. Bacteriol. 177 (7), 1915-1917 (1995)
REFERENCE 95204368

AUTHORS 3 (sites)

TITLE Futo, S., Seto, Y., Okada, M., Sato, S., Suzuki, T., Kawai, K., Imada, Y.
Recombinant 46-kilodalton surface antigen (P46) of Mycoplasma
hyopneumoniae expressed in Escherichia coli can be used for early
specific diagnosis of mycoplasmal pneumonia of swine by
enzyme-linked immunosorbent assay

J. Clin. Microbiol. 33 (3), 680-683 (1995)
MEDLINE 95270693

FEATURES Location/Qualifiers

Source 1. 1740

/organism="Mycoplasma hyopneumoniae"

/strain="ATCC25934 (strain J)"

/db_xref="taxon:2099"

-10_signal 187..192

misc_feature 198

/note="transcription initiation site"

RBS 222..230

CDS 235..1494

/codon_start=1

/transl_table=4

/product="46 kDa surface antigen"

/protein_id="BA04085.1"

/db_xref="PID:di004604"

/db_xref="PID:g1220114"

/db_xref="GI:1220114"

/translation="MKMLRKKFLYSIAIVATSLASIAFVAAGCGTSGSTSDSKP
OAEHLKHVNSDIRIALTDPNRPRISAGKDILSYDETEAITSITKNDQNNWL
TQANLSPAKGFIAPENGSGVGTAVNTIADRGIPVAYDRILTSIDKIDWYFSDN
EKVEQLGLSALGLGKEDGAFDSIDONNEYLKSHPQETISFYIAGSQDDNNQOY
FYANGKVLKMLNSQNKIIDLSPENAVYVPGWNYTAGRIQSFILTKINDPAGG
NKIKAVGSPASIFKGFPLNDGMAEOAITKLLEGFDOKIEFTVGDYNDKAKTFIK
DGQNMITYKPKDVLGVAVELRVLIKKKNSRSEVENELAKLNPISFKYDNTY
KVQGNLNTILVSPVIVTKANVNPDA"

BASE COUNT 670 a 239 c 258 g 373 t

ORIGIN

Query Match 82.9% Score 1476.4; DB 1; Length 1740;
Best Local Similarity 99.3%; Pred. No. 2.1e-140;
Matches 1483; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 289 ttacattttttatacaatttttaaaattactcttttaattatagttattttttatt 348
Db 1 TTTACATTTTTTATAACAATTTTAAATACCCCTTAAATTTATAGTATTTTTATT 60
Qy 349 ttttagtctaaattataaaattactgtgaattttattttgaattttataattgtacta 408
Db 61 TTTTAGTCTAAATATAAAATATCTTGAATTTTATTGAAATTTTATTAATTTAGTACTA 120
Qy 409 aaaaatacaaatattttctctattcttaagaaaaattcatttttttaaaaaattgattt 468
Db 121 AAAAATACAATATATTTTCTATCTTAAGAAAAATCAATTTTTTAAAAAATTTGATT 180
Qy 469 ttatagataattttgttataattgaattaaacttattgttgaagggaacaaatgaaa 528
Db 181 TTATAGTATAATTTATTGTATATTAATGAATTAACCTTGATTTGAAAGGGAACAAATGAA 240
Qy 529 aaatgtctagaaaaaaattctgtattcatcagctattttatgcaacttgcgtcatca 588
Db 241 AAAATGCTTAGAAAAAATTTCTTATTCATCAGCTATTTATGCAACTTCGTTGCATCA 300
Qy 589 attattcattgttcagcaggtgtgtggacagacagacaataacagttcaacttcgattct 648
Db 301 ATTATTGCAATTTGTTGAGCAGGTTGTGGACAGACAGAAATCAGGTTTCAGATTCT 360
Qy 649 aaacacacgagcagcgtctaaacataaaagtaagtaattgttcatttcgaatagcacta 708
Db 361 AAACCAAGCGGAGACTCTTAAACATAAAAGTAAGTAATGATTCTATTCGAATAGCACTA 420

Qy 709 accgateggataatctctogtagaattagtgcccaaaaagattattttttttgtgtat 768
Db 421 ACCGATCCGGATAATCTCTCGATTAATAGTAGTCCCAAAAAGATATTATTCTTATGTTAT 480
Qy 769 gaacagagcagcaacttcaaatcaaaaacaggagtcacaaaataactgactc 828
Db 481 GAAACAGAGGCGACAACTTCAACAATTAACAAAACAGGATGCACAAAATAACTGACTC 540
Qy 829 actcagcaagctaatttaagccagcgccaaaaggatttattattgccccctgaaaatgga 888
Db 541 ACTCAGCAAGCTAATTAAAGTCCAGCGCCAAAGAGATTATTATTGCCCTGAAATGGA 600
Qy 889 agtggagttggaactgctgtaatacaaatgctgataaagggaattccgattgttgcctat 948
Db 601 AGTGGAGTTGGAACTGCTGTTAATAACAATTCGTAAGAGAAATCCGATTGTTGCCAT 660
Qy 949 gatcgactaattactgactgataaataatgattgattgtttcttttggataatgaaaa 1008
Db 661 GATCGACTAATTACTGATCTGATAAATATGATTGGTATTGTTCTTTTGTAAATGAAAA 720
Qy 1009 gttgttgaaattacaaggtctttcacttgcgtcggtggtctattaggaaaaagaagtggct 1068
Db 721 GTTGGCGAATTACAGGTCTTTCACCTTGGCGGGTCTATTAGAAAAGAGATGGTCT 780
Qy 1069 tttgattcaattgatcaaatgaatgaatataatcaaatcaaatgccccaaagacaatt 1128
Db 781 TTTGATTCAATTGATCAAAATGAATATCTAAATTCATATGATATGCCCCAAGAGACAAT 840
Qy 1129 tctttttatacaatccgggttcccgaatgataaattcccatttttttaatttgggt 1188
Db 841 TCITTTTATCAATCGGGGTCCCAAGATGAATAATTTCCCAATATTTTATATATGGT 900
Qy 1189 gcaatgaagttacttaagaattcaatgaaaaaatcgcacaaataaaataattgtattatct 1248
Db 901 GCAATGAAGTACTTAAAGAAATTAAGAAAAATTCGCANAAATAAATTAATGATTATCT 960
Qy 1249 cctgaaggcgaaaaatgctgtttatgtccaggatgaattatggaactgcgcgtcaaga 1308
Db 961 CCTGAAGGCGAAAAATGCTGTTTATGTCAGGATGAAATTTATGAACTGCGGTCAAGA 1020
Qy 1309 atccaatttttcaacaattacaagaatccagcaggtggttaataatacaagtgtt 1368
Db 1021 ATCCAATCTTTTACAAATTAACAAGATCCAGCAGGTGGTAAATAAATCAAAAGCTGT 1080
Qy 1369 ggttcaaaacacagcttctatttcaaggatttcttgcccccaaatgatggaatggccgaa 1428
Db 1081 GGTCAAAACACAGCTTCTATTTTCAAGGATTTCTTGCCCCCAATGATGGAATGCCGAA 1140
Qy 1429 caagcaatcaccacaaattaaacttgaagggtttgtatcccaaaaaatcctttgtaactcgt 1488
Db 1141 CAAGCAATCACCAAATTAACCTTGAAGGATTTGATACCCCAAAAAATCTTTGTAACGTG 1200
Qy 1489 caagattataatgaagaacccaaactttatcaagagcggcgatcaaaatagacaatt 1348
Db 1201 CAAGATTATTAATGAAGCCAAACTTTTATCAAGACGGCGATCAAAATATGACAAT 1260
Qy 1549 tataaacctgataaagtttttaggaaaaagttgctgttgaaagttcttcgggttttaattgca 1608
Db 1261 TATAAACCTGTATAAGTTTATAGAAAAGTTGCTGTGAAAGTTCTTCGGGTTTAAATGCA 1320
Qy 1609 aagaaaaataagcactctagatcagaagtcgaaaaacgaactaaaaaactaccacaaat 1668
Db 1321 AAGAAAAATAAGCATCTAGATCAGAAGTCGAAAAGCAACTAAAGCAAAAATACCAAAAT 1380
Qy 1669 atttcatttaaatatataatcaaatataaagtcaaggttaaaataattatacaatt 1728
Db 1381 ATTTCATTAAATATGATAATCAACATATAAAGTCAAGGTGAAAAATATTATATACAAT 1440
Qy 1729 tttagtaagtcagtaattgttacaagaagctaaattgttgataaactcctgatccctaa 1782
Db 1441 TTAGTAAGTCCAGTAATTGTTACAAAGCTAAGGTGATGAATAATCTCTGATGCCATA 1494

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 1999, 13:07:05 ; Search time 473.11 Seconds
(without alignments)
11978.845 Million cell updates/sec

Title: US-08-913-430-1

Perfect score: 1782
Sequence: 1 atgaaaaaatgccactata.....ttgataatctgatgctaa 1782

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.*

1: gb_ba1.*
2: gb_ba2.*
3: gb_cm.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_st.*
14: gb_sts.*
15: gb_sy.*
16: gb_un.*
17: gb_vi.*
18: em_fun.*
19: em_htg.*
20: em_hum1.*
21: em_hum2.*
22: em_in.*
23: em_om.*
24: em_or.*
25: em_ov.*
26: em_pat.*
27: em_ph.*
28: em_pl.*
29: em_ro.*
30: em_sts.*
31: em_sy.*
32: em_un.*
33: em_vi.*
34: gb_htg1.*
35: gb_htg2.*
36: gb_in1.*
37: gb_in2.*
38: em_ba1.*
39: em_ba2.*
40: em_hum3.*
41: em_hum4.*
42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1476.4	82.9	1740	1	MYC46KDSA		D16682 Mycoplasma

ALIGNMENTS

RESULT 1

MYC46KDSA

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

2	1245.6	69.9	1260	5	E04787
3	835	46.9	843	5	E02476
4	392.8	22.0	396	5	E02474
5	356.8	20.0	360	5	E02475
6	99.8	5.6	153098	36	PFMAL3P2
7	98.6	5.5	2426	7	SD049822
8	97.4	5.5	219200	35	AC006279
9	97.4	5.5	4601	36	DMU11584
10	97.4	5.5	19517	37	DMU37541
11	97	5.4	65691	36	PFMAL3P1
12	96.6	5.4	12029	37	AE001426
13	95.8	5.4	15421	36	PFCOMP1RA
14	95.6	5.4	4601	36	DMU11584
15	95.4	5.4	19517	37	DMU37541
16	95.4	5.4	187544	35	AC004157
17	95	5.3	104992	35	AC005504
18	94.6	5.3	13741	36	CEW10C6
19	94.4	5.3	256172	35	AC005139
20	93.8	5.3	88295	11	AC004612
21	93.2	5.2	220475	35	AC006278
22	93.2	5.2	231699	35	AC006281
23	92.8	5.2	14867	37	AE001398
24	92.4	5.2	11829	37	AE001425
25	91.8	5.2	2243	36	PFVAR23A
26	91.6	5.1	14867	37	AE001398
27	91.6	5.1	37952	37	CELH28G03
28	91.4	5.1	47558	37	AF030694
29	91.2	5.1	2858	7	YSCMTG18
30	91.2	5.1	15363	37	MSQWTCG
31	91	5.1	167171	35	AC006269
32	90.4	5.1	21136	34	AC006827
33	90.2	5.1	242893	34	CEY53C12
34	90.2	5.1	1496	36	CEY53C12D
35	90	5.1	2000	7	CPMTATP6
36	90	5.1	2236	7	DPMTTRND
37	89.8	5.0	187544	35	AC004157
38	89.8	5.0	104992	35	AC005504
39	89.8	5.0	16019	36	MIDYTRN
40	89.8	5.0	2290	36	MIDYTRN
41	89.6	5.0	193188	34	AC006884
42	89.6	5.0	59306	42	AC007423
43	89.4	5.0	242893	34	CEY53C12
44	89.4	5.0	16343	36	AMFGENOM
45	89.2	5.0	95506	11	AC002407

MYC46KDSA 1740 bp DNA BCT 04-FEB-1999
Mycoplasma hyopneumoniae gene for 46 kDa surface antigen, complete cds.
D16682
g517134
D16682.1 GI:517134
46 kDa surface antigen; P46; lipoprotein.
Mycoplasma hyopneumoniae (Strain:ATCC25934 (strain J)) DNA.
Mycoplasma hyopneumoniae
Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
Mycoplasmataceae; Mycoplasma.
1 (bases 1 to 1740)
Futo,S
Direct Submission
Submitted (15-JUL-1993) to the DBJ/EMBL/GenBank databases. Satoshi Futo, Central Laboratory, Nippon Flour Mills Co., Ltd., Bioscience; 2114-2 Nukumizu, Atsugi, Kanagawa 243, Japan (Tel:0462-23-5115, Fax:0462-21-4970)
2 (bases 1 to 1620)
Futo,S., Seto,Y., Mitsuuse,S., Mori,Y., Suzuki,T. and Kawai,K.
Molecular cloning of a 46-kilodalton surface antigen (P46) gene from Mycoplasma hyopneumoniae: direct evidence of CGG codon usage

QY 272 ttctcataactctaaatcttcaatcttttt-tatacaaatcttttataaaatctactcttaat 330
 Db 1128 TTTTGTATTAAGATATATCTAAATCTGTTATATTTTAAATTTTGTATATAAAATTTAAAT 1069
 QY 331 ttatagatt 390
 Db 1068 TAATAATATTTAAATTTGAATATATAAATCTTTTAAATTTTCTGGAATATATTTTAAAT 1009
 QY 391 ttttataatttagtactaaatacaaatattt---ttttctattcttaagaaaaatttc 446
 Db 1008 TATTATATCATATATATATATATGCAATCTTCTAAATTAACATAATTTTAAATATAT 949
 QY 447 atttttataaaatatt 500
 Db 948 TAATAATATTAACATATCTCTGTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 889
 QY 501 acttgatttgaagggaacaaataagaaaaatgcttagaaaaaatcttctgtattc 560
 Db 888 TATTATTAATAATATATCAAAATTTAGTAATCCATTTTGAATAATTTTAAATTTTAAATTTTAAAT 829
 QY 561 agctattatgcaacttcgctgcatcaattattgcatctttgttcgagcaggttggtgaca 620
 Db 828 ATTGATT---AATTTTGTGAAAAATATATCAATTAATAAATAAATAAATGTCGATA 774
 QY 621 gacagaatca--ggttcaactctgattcttaacacacagccgagcgcgttaaaacataaa 678
 Db 773 GAATTAATAATTTGGTCTATGTTTCTGCAAAACACAGAAATTTGATGCGAAAAATTAAT 714
 QY 679 gtaagtaagtattcttctgaatagcacttaaccgcatccggaataatccctcgatgaattag 738
 Db 713 AGATATATTTTATCTAAATCAAAATTTGTAATTAATTAACATCAATTAATAGATAATAGATT 654
 QY 739 gcccaaaaaagattattttctttatg 763
 Db 653 ATAAATAAATAATATAAATCATG 629

RESULT 13
 V14508/c
 ID V14508 standard; DNA; 1511 BP.
 AC V14508;
 DT 20-MAY-1998 (first entry)
 DE AnePV entomopoxvirus thymidine kinase gene.
 KW Entomopoxvirus; spheroindin gene; AnePV; thymidine kinase; promoter;
 OS insect control; viral vaccine; ds.
 FS Amsacta moorei.
 FH Key
 FT Location/Qualifiers
 FT CDS
 FT complement (18.218)
 FT /*tag= a
 FT /product= Q1 open reading frame protein (see W41303)
 FT complement (234..782)
 FT /*tag= b
 FT /product= Q2 open reading frame protein (see W41304)
 FT 852..1511
 FT /*tag= c
 FT /product= Q3 open reading frame protein (see W41305)
 US5721352-A.
 PD 24-FEB-1998.
 PF 22-NOV-1993; 107755.
 PR 22-NOV-1993; US-107755.
 PR 19-FEB-1991; US-657584.
 PR 30-JAN-1992; US-827685.
 PR 12-FEB-1992; WO-U00855.
 PA (UYFL) UNIV FLORIDA RES FOUND.
 PI Grudl ME, Halli RL, Moyer RW;
 DR WPI; 98-168476/15.
 DR P-PSDB; W41303, W41304, W41305.
 DR New Entomopoxvirus nucleic acid sequences - used in DNA constructs
 PT and vectors for expression of heterologous genes in, e.g. insect
 PT cells
 PS Disclosure: Columns 51-54; 55pp; English.
 CC This sequence represents the Amsacta moorei entomopoxvirus (AnePV)
 CC thymidine kinase gene. EPV spheroindin and thymidine kinase promoters can

CC be used in DNA constructs and vectors for expression of heterologous
 CC genes in insects or mammalian cells, e.g. vectors containing Bacillus
 CC thuringiensis toxin genes for use in insect control, or recombinant
 CC vaccinia or swinepox viruses for use as viral vaccines.
 SQ Sequence 1511 BP; 640 A; 128 C; 98 G; 645 T;

Query Match 4.3%; Score 77.4; DB 1; Length 1511;
 Best Local Similarity 49.8%; Pred. No. 0.013;
 Matches 341; Conservative 0; Mismatches 326; Indels 18; Gaps 5;

QY 92 tegtacattaaagaagtagaacaagaagttattatgtataacacattagcgaatccttaagaa 151
 Db 1308 TATTTAACTCAACTACTACTAATAGGAAAAACATTTATTAAGTTACCAATTTATTTTAG 1249
 QY 152 aaaaataaagttttctctatt 211
 Db 1248 ATATTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 1189
 QY 212 agtattatcaagtcggtatt 271
 Db 1188 ATGATTCACAATTTAAATTTTCTATAGAATGGGTAGTATATATTTACTTATATTCCTAT 1129
 QY 272 ttccataatctaaatatt 330
 Db 1128 TTTTGTATTAAGATATATCTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 1069
 QY 331 ttatagatt 390
 Db 1068 TAATAATATTTAAATTTGAATATATAAATCTTTTAAATTTTCTGGAATATATTTTAAAT 1009
 QY 391 ttttataatttagtactaaatacaaatattt---ttttctattcttaagaaaaatttc 446
 Db 1008 TATTATATCATATATATATATATGCAATCTTCTAAATTAACATAATTTTAAATATAT 949
 QY 447 atttttataaaatatt 500
 Db 948 TAATAATATTAACATATCTCTGTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 889
 QY 501 acttgatttgaagggaacaaataagaaaaatgcttagaaaaaatcttctgtattc 560
 Db 888 TATTATTAATAATATATCAAAATTTAGTAATCCATTTTGAATAATTTTAAATTTTAAATTTTAAAT 829
 QY 561 agctattatgcaacttcgctgcatcaattattgcatctttgttcgagcaggttggtgaca 620
 Db 828 ATTGATT---AATTTTGTGAAAAATATATCAATTAATAAATAAATAAATGTCGATA 774
 QY 621 gacagaatca--ggttcaactctgattcttaacacacagccgagcgcgttaaaacataaa 678
 Db 773 GAATTAATAATTTGGTCTATGTTTCTGCAAAACACAGAAATTTGATGCGAAAAATTAAT 714
 QY 679 gtaagtaagtattcttctgaatagcacttaaccgcatccggaataatccctcgatgaattag 738
 Db 713 AGATATATTTTATCTAAATCAAAATTTGTAATTAATTAACATCAATTAATAGATAATAGATT 654
 QY 739 gcccaaaaaagattattttctttatg 763
 Db 653 ATAAATAAATAATATAAATCATG 629

RESULT 14
 Q241134/c
 ID Q241134 standard; DNA; 1671 BP.
 AC Q241134;
 DT 09-NOV-1992 (first entry)
 DE 50 KD subunit of Scler.
 KW Endonuclease Scler; PAGE; chromatography; ENS2 gene; ss.
 OS Sacchromyces cerevisiae.
 FH Key
 FT Location/Qualifiers
 FT cds
 FT 1..1589
 FT /*tag= a
 FT /label= Scler_50kd_subunit
 FT 1..81
 FT exon

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 1999, 11:59:45 ; Search time 288.99 Seconds
(without alignments)
12163.240 Million cell updates/sec

Title: US-08-913-430-1
Perfect score: 1782
Sequence: 1 atgaaaaaatgccactata.....ttgataatcctgatgcctaa 1782

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*

- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	82.4	4.6	688	45	AU001063	AU001063 AU001063
2	82.2	4.6	700	45	AU006427	AU006427 AU006427
3	81.4	4.6	717	45	AU000694	AU000694 AU000694
4	80.2	4.5	736	45	AU001091	AU001091 AU001091
5	80	4.5	527	44	AU037422	AU037422 AU037422
6	79.4	4.5	593	45	AU002045	AU002045 AU002045
7	79.4	4.5	633	45	AU004482	AU004482 AU004482
8	78.2	4.4	507	27	W82081	W82081 me96a06.r1
9	78	4.4	471	46	AI404883	AI404883 GH24733.5
10	77.8	4.4	690	49	AI619044	AI619044 AEMTBG38
11	77.8	4.4	699	49	AI621573	AI621573 AEMTAN90
12	77.8	4.4	705	49	AI648282	AI648282 AEMTAA94
13	77.8	4.4	723	49	AI650127	AI650127 AEMTAO28
14	77.8	4.4	654	49	AI650135	AI650135 AEMTA095
15	77.8	4.4	707	49	AI650165	AI650165 AEMTAR49
16	77.8	4.4	699	49	AI658356	AI658356 AEMTBF79
17	77.8	4.4	711	49	AI658398	AI658398 AEMTBK45
18	77.8	4.4	724	49	AI658446	AI658446 AEMTBN64
19	77.4	4.3	700	48	AI557583	AI557583 pt2.1-1.D
20	77.2	4.3	849	45	AU005735	AU005735 AU005735
21	77.2	4.3	679	49	AI648352	AI648352 AEMTAE89
22	76.8	4.3	776	47	AI525944	AI525944 DU145-1.G
23	76.8	4.3	711	49	AI657490	AI657490 AEMTBB16
24	76.2	4.3	735	49	AI650096	AI650096 AEMTAM27
25	76.2	4.3	727	49	AI650129	AI650129 AEMTA040
26	76	4.3	481	44	AU038892	AU038892 AU038892
27	76	4.3	479	45	AI396905	AI396905 fb16a11.y
28	74.8	4.2	573	45	AU001034	AU001034 AU001034
29	74.6	4.2	3707	39	C83838	C83838 C83838 D1ct
30	74.6	4.2	702	49	AI648316	AI648316 AEMTAC90
31	74	4.2	602	51	AI723414	AI723414 hcgl544.T
32	73.4	4.1	479	25	W08634	W08634 mb42b12.r1
33	73.2	4.1	695	49	AI650099	AI650099 AEMTAM46
34	73	4.1	604	51	AI723413	AI723413 hcgl544.T
35	72.6	4.1	273	39	C84830	C84830 C84830 D1ct
36	72.2	4.1	624	35	AA550217	AA550217 1335m3 gm
37	72	4.0	613	44	AI293736	AI293736 LP06939.5
38	72	4.0	684	44	AI294319	AI294319 LP07659.5
39	72	4.0	321	44	AU037255	AU037255 AU037255
40	72	4.0	559	45	AU005020	AU005020 AU005020
41	71.6	4.0	849	45	AU005735	AU005735 AU005735
42	71.6	4.0	562	49	AI648323	AI648323 AEMTAD25
43	71.4	4.0	705	43	AI238819	AI238819 GH14939.5
44	71.4	4.0	624	48	AI537728	AI537728 Plitest.B
45	71.4	4.0	256	50	AU060591	AU060591 AU060591

ALIGNMENTS

RESULT 1
AU001063 AU001063 688 bp mRNA EST 15-JAN-1999
LOCUS AU001063 Bombyx mori p50(Daizo) Bombyx mori cDNA clone fb11133f,
DEFINITION mRNA sequence.
ACCESSION AU001063
NID 94157307
VERSION AU001063.1 GI:4157307

KEYWORDS

SOURCE EST.
ORGANISM domestic silkworm.

LOCUS Bombyx mori
DEFINITION Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Bombycidae; Bombyx.

REFERENCE 1 (bases 1 to 688)
AUTHORS Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Establishment of cDNA database of Bombyx mori
JOURNAL Unpublished (1999)
COMMENT On Oct 8, 1998 this sequence version replaced gi:3726239.

Contact: Mita K

Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
PROJECT = 'CREST project by JST'

PROJCT = 'CREST project by JST'

Location/Qualifiers

1. .688

/organism="Bombyx mori"

/strain="p50(Daizo)"

/db_xref="taxon:7091"

/map="8"

/clone="fbf1133f"

/clone_lib="Bombyx mori p50(Daizo)"

285 a 33 c 55 g 315 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 4.6%; Score 82.4; DB 45; Length 688;

Matches 290; Conservative 0; Mismatches 271; Indels 15; Gaps 3;

Qy 46 taattaaataacattttcttcattgcccagaaatttttaagaatttagtacattaaaaa 105

Db 70 TATTAAATAATTAATTTATTTTAAAGTTTCGCAATTTAAAGAGTTAATTAATTAATA- 129

Qy 106 gtagaacaaagtattataatgtaaacattagcgcgaatccttaagaacaaattaaagt 165

Db 129 -----ATAGTTTGTAGTATAATATTTTAAATAATTAATTAATTAATTAATTAATTA 181

Qy 166 tatctatttttttaacgaaatccaccagcgaatttttaagaatttagtacattaaaaa 225

Db 182 TTCGTTTAAATATCTAGTTTATTTTAAAGAGTTTAAATTTTAAATTTTAAATTTTAAATTT 241

Qy 226 cgtatttttttctactaataattattggaatttgcattttccataactaa 285

Db 242 TTTTATTAATTAATTAATTAATTAATTTTAAATAATTAATTAATTAATTAATTAATTTT 301

Qy 286 aattttacattttttatacaatttttaaaaaatta--ctcttaatttagtatttt 343

Db 302 AATTAATAATTTTAAATAATTTTAAATAATTAATTAATTAATTAATTAATTTTAAATTT 361

Qy 344 ttatttttagtctaaattataaaattattctggaatttttgaatttttataatttag 403

Db 362 AATTTTATTTTATTAATAATTTTAAATAATTAATTAATTAATTAATTTTAAATTTTAAATTT 421

Qy 404 tactaaaaatacaaatatttttctactaataattattggaatttttgaatttttgaatttt 463

Db 422 TATATAAATAATTTTAAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTT 481

Qy 464 gatttttagta-----taattgtttgataattggaatttttgaatttttgaattttgaa 518

Db 482 ATTAATAATAATTTTAAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTT 541

Qy 519 caaaatgaaaaaatcgttagaaaaattcttctgatttctatcattcagctatttagcaacttc 578

Db 542 TTATATTCACCTGTTTATCAAAACATGCTCTTTTGTGTAATTAATTTTAAAGCTCTAACTGC 601

Qy 579 gcttgcaatcattgctattgttgcagcaggttg 614

Db 602 CCACTGATATATTTTAAAGGCTGAGTATTTTG 637

RESULT 2

LOCUS AU006427

DEFINITION AU006427 Bombyx mori p50(Daizo) Bombyx mori cDNA clone wv41020,

mRNA sequence.

ACCESSION AU006427

NID 94163811

VERSION AU006427.1 GI:4163811

KEYWORDS EST.

SOURCE domestic silkworm.

ORGANISM Bombyx mori

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;

Ditrysia; Bombycoidea; Bombycidae; Bombyx.

REFERENCE 1 (bases 1 to 700)

AUTHORS Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.

TITLE Establishment of cDNA database of Bombyx mori

JOURNAL Unpublished (1999)

COMMENT On Jan 17, 1998 this sequence version replaced gi:2043379.

Contact: Mita K

Genome Research Group

National Institute of Radiological Sciences

Anagawa 4-9-1, Inage, Chiba 263-8555, Japan

Email: kmita@nirs.go.jp

PROJECT = 'CREST project by JST'

Location/Qualifiers

1. .700

/organism="Bombyx mori"

/strain="p50(Daizo)"

/db_xref="taxon:7091"

/clone="wv41020"

/clone_lib="Bombyx mori p50(Daizo)"

294 a 25 c 59 g 322 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 4.6%; Score 82.2; DB 45; Length 700;

Matches 290; Conservative 0; Mismatches 278; Indels 11; Gaps 3;

Qy 46 taattaaataacattttcttcattgcccagaaatttttaagaatttagtacattaaaaa 105

Db 122 TATTAAATAATTTATTTTAAAGTTTCGCAATTTAAAGAGTTTAAATTTTAAATTTTAAATTT 181

Qy 106 gtagaacaaagtattataatgtaaacattagcgcgaatccttaagaacaaattaaagt 165

Db 181 -----ATAAGTTTGTAGTATAATATTTTAAATAATTAATTTTAAATAATTTTAAATAATTT 233

Qy 166 tatctatttttttaacgaaatccaccagcgaatccttgcagatttatcaagt 225

Db 234 TTCGTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTT 293

Qy 226 cgtatttttttctactaataattattggaattttgcaattttccataactaa 285

Db 294 TTTTATTAATTAATTAATTAATTAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTT 353

Qy 286 aattttacattttttatacaatttttaaaaaatta--ctcttaatttagtatttt 343

Db 354 AAATTAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTT 413

Qy 344 ttatttttagtctaaattataaaattattctggaatttttgaatttttataatttag 403

Db 414 AAATTTTATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTT 473

Qy 404 tactaaaaatacaaatatttttctactaataattattggaatttttgaatttttgaatttttgaatttt 463

Db 474 TATATAAATAATTTTAAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTT 533

Qy 464 gat-ttttagtataaattttgttataattggaattttggaattttggaattttggaattttggaatttt 522

```

Db 534 ATTAATTAATAAATAATTAATTAATTAATTAATAAATAAAGGAATTCGCCAAATAT 593
QY 523 atgaaaaaattcgttagaaaaaattctgttattcatcagcattattgaacttcgctt 582
Db 594 TTATATTCACATGTTTATCAAAAACATGCTCTTTGGTAATAATTAAGTCATATCGCC 653
QY 583 gcatcaattattgatttctgttgagcaggggtgtgacag 621
Db 654 CACTGATATATTTAATAAGGCTGCAGTATTTTGACTG 692

RESULT 3
LOCUS AU000694 717 bp mRNA EST 15-JAN-1999
DEFINITION AU000694 Bombyx mori p50(Daizo) Bombyx mori cDNA clone e40883, mRNA
sequence.
ACCESSION AU000694
NID 94156938
VERSION AU000694.1 GI:4156938
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
TITLE Ditrysia; Bombycoidea; Bombycidae; Bombyx.
JOURNAL Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
COMMENT Establishment of cDNA database of Bombyx mori
Unpublished (1999)
On Oct 6, 1998 this sequence version replaced gi:3709815.

Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@exs64.nirs.go.jp
PROJECT - 'CREST project by JST'

FEATURES
source
1. .717
/organism="Bombyx mori"
/strain="p50(Daizo)"
/db_xref="taxon:7091"
/clone="e40883"
BASE COUNT 304 a 21 c 59 g 333 t
ORIGIN

Query Match 4.6%; Score 81.4; DB 45; Length 717;
Best Local Similarity 51.2%; Pred. No. 0.0054;
Matches 270; Conservative 0; Mismatches 246; Indels 11; Gaps 3;

QY 46 taataaaattacattttcttcatttgcgcagaaattttaagaatttagtacattaaaaa 105
Db 168 TATTAAAAATAATTTTATTTTAAAGTTTCGGAATTTAAAGAGTTAATTAATTAATA- 227
QY 106 gtagaacaaaggatttaattgaataacattgcgcgaattcttaagaacaaattaaaagttt 165
Db 227 -----ATAGTTATTGTAGTATAATAATTTTAAATAATTAATTTGAATGAATGTTA 279
QY 166 tatctatttttttaactgaataccaccaggcataactcttgcagttatttatcaagt 225
Db 280 TTCGTTTTTAATATATCTAGTTTTTTATAGAAAAAATTTAATTTTATTAAATAGATT 339
QY 226 cggatttttttcatattctactaaaattatttgaatttgatttccattccataatactaa 285
Db 340 TTTTATTAAATTAATAAATAATATATTTTAAATAAATAATTTAAGGATAATCTTT 399
QY 286 aattttacatttttttaacaaattttaaaatta--ctcttaatttagtattttt 343
Db 400 AAATTAATAATTTTATATAATTTTATTAATAAAGAAATTTATATAATTAATTAAT 459
QY 344 ttatttttagctaaattataaaattatcttgcgaattttatttgaatttttataattag 403

```

```

Db 460 AAATTTTATTTTATTAATAAATAATTTTACTAATAGAAAAATTTATTTATTTT 519
QY 404 tactaaaaatacaataatttttcttcttcttaagaacaaattcattttttaaataatt 463
Db 520 TATATAAATAAATAATTTTAAATTAATAAATAATTAATTAATGAATAAATAGTAAATTT 579
QY 464 gat-ttttatgtataatttggttgtatataatgaataactgtattgaaagggaacaa 522
Db 580 ATTAATTAATAAATAATTTAATTAATTTAATTAATAAATAAAGGAATTCGGCAATAT 639
QY 523 atgaaaaaattcgttagaaaaaattctgttattcatcagctattta 569
Db 640 TTATATTCACATGTTTATCAAAAACATGCTCTTTTGTATAATAATTA 686

RESULT 4
LOCUS AU001091 736 bp mRNA EST 15-JAN-1999
DEFINITION AU001091 Bombyx mori p50(Daizo) Bombyx mori cDNA clone fbm0036f,
mRNA sequence.
ACCESSION AU001091
NID 94157335
VERSION AU001091.1 GI:4157335
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
TITLE Ditrysia; Bombycoidea; Bombycidae; Bombyx.
JOURNAL Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
COMMENT Establishment of cDNA database of Bombyx mori
Unpublished (1999)
On Oct 8, 1998 this sequence version replaced gi:3726544.

Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@exs64.nirs.go.jp
PROJECT - 'CREST project by JST'

FEATURES
source
1. .736
/organism="Bombyx mori"
/strain="p50(Daizo)"
/db_xref="taxon:7091"
/map="911F06; 13; 13q14.12-13q14.3; 8"
/clone="fbm0036f"
/clone_lib="Bombyx mori p50(Daizo)"
BASE COUNT 311 a 22 c 59 g 343 t 1 others
ORIGIN

Query Match 4.5%; Score 80.2; DB 45; Length 736;
Best Local Similarity 49.8%; Pred. No. 0.0078;
Matches 262; Conservative 0; Mismatches 254; Indels 10; Gaps 2;

QY 46 taattaaaaattacattttcttcttgcgcagaaattttaagaatttagtacattaaaaa 105
Db 187 TATTAATAAATAATTTTATTTTAAAGTTTCTCGAATTTAAAGAGTTAATTAATTAATA- 246
QY 106 gtagaacaaaggatttaattgaataacattgcgcgaattcttgaacaaataaaggttt 165
Db 246 -----ATAGTTATTGTAGTATAATAATTTTAAATAATTAATTTGAATGAATGTTA 298
QY 166 tatctattttttttaaactgaataccaccaggcataactcttgcagttatttatcaagt 225
Db 299 TTCGTTTTTAATATATCTAGTTTTTTATAGAAAAAATTTAATTTTATTAAATAGATT 358
QY 226 cggatttttttcatattctactaaaattatttgaatttgatttccattccataactaa 285
Db 359 TTTTATTAAATTAATAAATAATATATTTTAAATAAATAATTTAAGGATAATCTTT 418

```



```

Environmental Toxicology
University of California, Riverside
5419 Boyce Hall, Riverside, CA 92521, USA
Tel: 909 787 3547
Fax: 909 787 3087
Email: sarjeet.gill@ucr.edu
Seq primer: CCAGCTCTAATCGACTCACTAT
High quality sequence stop: 600.
Location/Qualifiers
1. 690
/organism="Aedes aegypti"
/clone="BC38"
/db_xref="taxon:7159"
/sex="female"
/tissue_type="malpighian tubules and gut"
/dev_stage="adult"
/lab_host="DH10B"
/notes="vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
cDNA was cloned into the SalI/NotI sites of pSPORT1."
293 a 25 c 58 g 314 t
BASE COUNT
ORIGIN

```

Query Match	4.43;	Score 77.8;	DB 49;	Length 690;
Best Local Similarity	47.9%;	Pred. No. 0.017;		
Matches 259;	Conservative 0;	Mismatches 277;	Indels 5;	Gaps 1;
QY	35	agataataaataaataacatttccttcattgcccagaaattttaagaattag	94	
Db	42	AGAAATAAATAATTTAATAATAGATATATTGTATGTTAAAGAAATAATGAAATAATTG	101	
QY	95	tacattaaaaagtagaacaaaagttattaatgtataacattagcgaatccttaagaataa	154	
Db	102	AAAAATTTTATTTAAAAAGAAAAATTAATTATTGTACCTGTGTATCAGGGTTTATTA	161	
QY	155	attaaaagtttatactattttttttaatcgaaatcccaaccaggcataaattotttgtcagt	214	
Db	162	ATTAATTAATTAATTAATTAATTTTCGGAATTTTAAAGATTTTAATTTATATATAAAAGT	221	
QY	215	attatcaagtcgggtatttttttcattttctactaaaattatttgaaatttgcatttt	274	
Db	222	TACGTGGAATAATTAATTTAAATATATAATAGAAATGAATGTAATCGTTTAAAAAT	281	
QY	275	ccataactaaaaatttacatttttttataacaatttttaaaaaattactcttttaattat	334	
Db	282	ATATCTAGTTTTTAAAGAAATAAATTAATTTAGATTTTATAAATTAATAAATTTTATTTAT	341	
QY	335	agattttttttatttttagctcaaaattataaaa-----ttactctgaattttattgaa	389	
Db	342	ATAGTATTTTTTATTATAAAATTAATTAATTTAGGGGATTAAGCTTTAAAAATAAATTTTA	401	
QY	390	ttttataatttagtactaaaaatacaaatatttttccattcttaagaataatcatt	449	
Db	402	AAATTTTAATAAAATATTTTAATAATGTAAGCTTAAAAATAGTTATCATTAATAATTG	461	
QY	450	ttttaaaaaaatgatttttattagataaatttgtttgtataaattgaattaaacttgattt	509	
Db	462	TTATTAATTTTATTTAAAAATTTTATTTATTTTGTGTAATTAATAATTTATTTTAAAAATAT	521	
QY	510	gaaggggaacaaaatgaaaaaatgcttagaaaaaatctctgtattcatcagctattta	569	
Db	522	AAATTTTAATGATTAATAATTTTATAAATATGATAAATAGTATATTTTAATTTATATAAA	581	
QY	570	t 570		
Db	582	t 582		

RESULT	11
AI621573	
LOCUS	AI621573
DEFINITION	AEMTAN90 Aedes aegypti MT psport Library Aedes aegypti cDNA clone
	699 bp mRNA EST 21-APR-1999

AN90 5', mRNA sequence.

ACCESSION AI621573
NID 94630699
VERSION AI621573.1 GI:4630699
KEYWORDS EST.
SOURCE yellow fever mosquito.
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Culicidae; Aedes.
1 (bases 1 to 699)
Gill, S.S., Ross, L.S. and Wadiak, H.
Expressed sequence tags of cDNA clones from an enriched Malpighian
tubule and gut library from Aedes aegypti
Unpublished (1999)
ON Mar 20, 1998 this sequence version replaced gi:2980567.
JOURNAL
COMMENT

Contact: Gill SS
Environmental Toxicology
University of California, Riverside
5419 Boyce Hall, Riverside, CA 92521, USA
Tel: 909 787 3547
Tel: 909 787 3547
Fax: 909 787 3087
Email: sarjeet.gill@ucr.edu
Seq primer: CCAAGCTCTAATACGACTCAGTAT
High quality sequence stop: 600.

```

FEATURES
source
Location/Qualifiers
1. .699
/organism="Aedes aegypti"
/db_xref="taxon:7159"
/map="21q22"
/clone="AN90"
/clone_lib="Aedes aegypti MT pSPORT Library"
/sex="female"
/tissue_type="malpighian tubules and gut"
/dev_stage="adult"
/lab_host="DH10B"
/note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
cdna was cloned into the SalI/NotI sites of pSPORT1."
300 a 25 c 56 g 318 t
BASE COUNT
ORIGIN

```

[illegible]

Db 400 AATTTTAAATAAATATTAAATAAGTAAAGCTTAAATAATAGTTATCATTAATAAATTTG 459

Qy 450 ttttaaaaaaattgattttttatagataaattgttctgataaattgaattgaattgattt 509

Db 460 TTATAAATTTTAAATTTTATTTATTTATTTGTTAAATTAATAATTTATTAATAATAT 519

Qy 510 gaaagggaacaaatgaaaaaattcttagaaaaaattcttctgattcatcagctattta 569

Db 520 AAATTTTAAATGATAAATTTTATAAATTTATGATAAATTTAGTATATTTTATATATAA 579

Qy 570 t 570

Db 580 T 580

RESULT 12

AI648282 AI648282 705 bp mRNA EST 30-APR-1999

LOCUS AEMTA094 Aedes aegypti MT pSPORT Library Aedes aegypti cDNA clone

DEFINITION AA94 5', mRNA sequence.

ACCESSION AI648282

NID 94726961

VERSION AI648282.1 GI:4726961

KEYWORDS EST.

SOURCE yellow fever mosquito.

ORGANISM Aedes aegypti

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae; Aedes.

AUTHORS Gill, S.S., Ross, L.S. and Wadiak, H.

TITLE Expressed sequence tags of cDNA clones from an enriched Malpighian tubule and gut library from Aedes aegypti

JOURNAL Unpublished (1999)

COMMENT On Jun 5, 1998 this sequence version replaced gi:3188103.

CONTACT: Gill SS

Environmental Toxicology

University of California, Riverside

5419 Boyce Hall, Riverside, CA 92521, USA

Tel: 909 787 3547

Fax: 909 787 3087

Email: sarjeet.gill@ucr.edu

Seq primer: CCAAGCTCTAATACGACTCCTAT

High quality sequence stop: 705.

FEATURES

Source

1..705

/organism="Aedes aegypti"

/db_xref="taxon:7159"

/clone="AA94"

/clone_lib="Aedes aegypti MT pSPORT Library"

/sex="female"

/tissue_type="malpighian tubules and gut"

/dev_stage="adult"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The cDNA was cloned into the SalI/NotI sites of pSPORT1."

BASE COUNT 297 a 26 c 55 g 327 t

ORIGIN

Query Match 4.4%; Score 77.8; DB 49; Length 705;

Best Local Similarity 47.9%; Pred. No. 0.017;

Matches 259; Conservative 0; Mismatches 277; Indels 5; Gaps 1;

Qy 35 agtatataaataataaataacattttcttcatttgcgcagaatttttaagaattag 94

Db 60 AGAATAAATAATTTAATAATAGTATATTTGTTATTTGTTAAAGAAATTTGAATAATTTG 119

Qy 95 tacattaaaaagtagaacaagaattattatgtaaacattgacgaatcccttaagaaaa 154

Db 120 AAAAATTTTATTTAAAGAAATTTTAAATTTATTTGTTACCTTGTGTATCAGGGTTTATTA 179

Qy 155 attaaagtttttctatttttttttaacgaaatccaacccaggcataaatactttgttcagt 214

Db 180 ATTAATTAATTAATTAATTAATTTTCTCGAATTTTAAAGATTTAATTAATTAATAAAGT 239

Qy 215 atttatacagtcggtatttttttctattttctactataaataattttgaatttgcattt 274

Db 240 TACTGTGGAATAATTAATTTTAAATATATAATTTAGAAATGAATGTTAAATCGTTTAAAT 299

Qy 275 ccaataactaaaaattttacatttttttatacaaattttaaaaaattactcttcaatttat 334

Db 300 ATATCTAGTTTTTAAAGAAATAATTTAATTTAGATTTAATAATTTAAATTTTATTTAT 359

Qy 335 agtattt 389

Db 360 ATAGTATTTTATTTTATAAATAATTAATTTTAAAGGATTAGCTTTAAATAAATTTTA 419

Qy 390 tttttatacatttgtagtactaaaaatacaaatattttttctctactataaagaataattc 449

Db 420 AATTTTAAATAAATAATTTTAAATGTAAGCTTAAATAATAGTATATTAATTTATATAA 479

Qy 450 ttttaaaaaaattgatt 509

Db 480 TTATAATTTTAAATTTTATTTTATTTTATTTTAAATTAATTTTATTTTAAATAAT 539

Qy 510 gaaagggaacaaatgaaaaaattgcttagaaaaaattcttctgattcatcagctattta 569

Db 540 AAATTTTAAATGATAAATAATTTAATAATTTATGATAAATAATTTATTTATATAA 599

Qy 570 t 570

Db 600 T 600

RESULT 13

AI650127

LOCUS AEMTA028 Aedes aegypti MT pSPORT Library Aedes aegypti cDNA clone

DEFINITION AO28 5', mRNA sequence.

ACCESSION AI650127

NID 94734106

VERSION AI650127.1 GI:4734106

KEYWORDS EST.

SOURCE yellow fever mosquito.

ORGANISM Aedes aegypti

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae; Aedes.

AUTHORS Gill, S.S., Ross, L.S. and Wadiak, H.

TITLE Expressed sequence tags of cDNA clones from an enriched Malpighian tubule and gut library from Aedes aegypti

JOURNAL Unpublished (1999)

COMMENT On Mar 10, 1998 this sequence version replaced gi:2949119.

CONTACT: Gill SS

Environmental Toxicology

University of California, Riverside

5419 Boyce Hall, Riverside, CA 92521, USA

Tel: 909 787 3547

Fax: 909 787 3087

Email: sarjeet.gill@ucr.edu

Seq primer: CCAAGCTCTAATACGACTCCTAT

High quality sequence stop: 723.

FEATURES

Source

1..723

/organism="Aedes aegypti"

/db_xref="taxon:7159"

/clone="AO28"

/clone_lib="Aedes aegypti MT pSPORT Library"

/sex="female"

/tissue_type="malpighian tubules and gut"

/dev_stage="adult"

/lab_host="DH10B"

```

/Note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
cdna was cloned into the SalI/NotI sites of pSPORT1."
302 a      32 c      65 g      324 t
BASE COUNT
ORIGIN

```

Query Match	4.4%	Score 77.8;	DB 49;	Length 723;
Best Local Similarity	47.9%;	Pred. No. 0.017;		
Matches 259;	Conservative	0;	Mismatches 277;	Indels 5;
Gaps	1;			

Qy 95 tacattaaaagtgaacaaagttatttaataatgaacatttagcgcaatccttaagaaaa 154
| | | | | | | | | | | | | | | | | |
Db 97 AAAAAATTTTATAAAAGACAAAATTAAATTATTGACCTTGTCATCAGGGTTATTA 156
| | | | | | | | | | | | | | | | | |

Qy 155 attaaaagtttattctatatttttttaatacgaatccaccaggcataaatctttgtcagt 214
||||| | | ||||| ||||| | | | | |
Db 157 ATTAAATTAAATTAATTAATTAATTTTTCCTCGAATTTTAAAGATTATAATATAATAAAGCT 216
||||| | | ||||| ||||| | | | | |

Qy 215 attatcaagtcggtatttttttcattattcttactaaatattatttgaatttgcatttt 274

Qy 275 ccataactaaaaatcttttacattttttttataacaaatttttaaaaaattactctttaatttat 334

DB 2// ATATCTAGTTTTTAAAGAAATAAATTAAATTAGATTATATAAATAAATAATTTATAT 335

Qy 335 agtatttttttttttttttagtctaaattataaaa-----ttatcttgaattttttatttgaa 389

Accession	Sequence	Length
Db	337 ATAGTATTTTATTTATATAATTTAAAGGGATTAGCTTTAAATAAATTTT	396
Qy	390 tttttataatttagtactaaaaaacacaaatattttttcctattcttaagaaaaattcatt	449

Db 397 AATTTTAAATAAATATTTTAAATAATGTAAGCTTAAAAATAGTATATCATTAATAAATTG 456

Qy 450 ttttaaaaaaaatttgatgttttttatagataaattgtttgtataaattgaaataaactgattt 509

Db 457 TTATAATTATTTTAAAAATTTTATTTATTTATGTTAAATTAAATTATTTATTAANAAT 516

Ov 510 caaagggacacaaaatgaaaaaaatgcttgaaaaaaaatcttgtattcatcagcattatta 569

517 AAATTTTATGATAATAATTTATATAATTATGATAAATAGTATATTATTTTATATAAA 576

RESULT	14
AI650135	.
LOCUS	AI650135 654 bp mRNA EST 04-MAY-1999
DEFINITION	AEMTAO95 Aedes aegypti MT PSPORT Library Aedes aegypti cDNA clone AO95 5', mRNA sequence.

ACCESSION	AI650135
NID	94734114
VERSION	AI650135.1
KEYWORDS	GI:4734114 EST.
SOURCE	yellow fever mosquito. todes acquati ORGANISM

Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 Culicoidae; Culicidae; Aedes.
 1 (bases 1 to 654)
 Gill, S.S., Ross, L.S. and Wadiak, H.
 Expressed sequence tags of cDNA clones from an enriched Malpighian
 tubule and gut library from *Aedes aegypti*
 Unpublished (1999)
 On Mar 10, 1998 this sequence version replaced g1:29491127.

Contact: Gill SS

RESULT	15
AI650165	
LOCUS	
DEFINITION	

EST 04-MAY-1999
y Aedes aegypti cDNA c

707 bp mRNA
s aegypti MT pS

AI650165
AEMTAR49

[illegible]04-MAY-1999
gypti cDNA cloneEST
y Ae

707 bp mRNA
s aegypti MT pS

AI650165
AEMTAR49

[illegible]

AR49 5', mRNA sequence.

ACCESSION AI650165
 NID G4734144
 VERSION AI650165.1 GI:4734144
 KEYWORDS EST.
 SOURCE yellow fever mosquito.
 ORGANISM Aedes aegypti

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae; Aedes.

AUTHORS Gill, S.S., Ross, L.S. and Wadiak, H.

TITLE Expressed sequence tags of cDNA clones from an enriched Malpighian tubule and gut library from Aedes aegypti

JOURNAL Unpublished (1999)

COMMENT On Mar 10, 1998 this sequence version replaced gi:2949158.

Contact: Gill SS

Environmental Toxicology

University of California, Riverside

5419 Boyce Hall, Riverside, CA 92521, USA

Tel: 909 787 3547

Fax: 909 787 3087

Email: sarjeet.gill@ucr.edu

Seq primer: CCAAGCTCTATACGACTCACTAT

High quality sequence stop: 707.

FEATURES

Location/Qualifiers

1..707

/organism="Aedes aegypti"

/db_xref="taxon:7159"

/clone="AR49"

/clone_lib="Aedes aegypti MT pSPORT Library"

/sex="female"

/tissue_type="malpighian tubules and gut"

/dev_stage="adult"

/lab_host="DH10B"

/note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; The cDNA was cloned into the SalI/NotI sites of pSPORT1."

BASE COUNT 304 a 26 c 56 g 321 t

ORIGIN

Query Match

Best Local Similarity 4.4%; Score 77.8; DB 49; Length 707;

Matches 259; Conservative 0; Mismatches 277; Indels 5; Gaps 1;

Qy 35 agtatataaaataataaaattacattttcttcatttgcgcagaatttttaagaattag 94

Db 47 AGAATAAATAATTTTAATAATAGTATATTTGTATTGTTAAAGAAATTTGAATAATTG 106

Qy 95 tacattaaagtagaacaaagttatttaatacgaatcgaacattagcgcacatccctaaagaaaa 154

Db 107 AAAAAATTTTATTTAAAGAAATTTAAATTTATTTGTTACCTTGTTGTCAGGGTTATTA 166

Qy 155 attaaagttttatctatttttttaatacgaatcccaaccggcagataatctttgtcagt 214

Db 167 ATTAATAATTAATTAATTAATTTTCGGAATTTTAAAGATTTAATTTATATATAAAGT 226

Qy 215 atttacaagtcggtatttttctattttctactataaaattatttgaatttcatttt 274

Db 227 TACTGTGGAATAATTTTAAATATATATAATTAAGAAATGAAATGTTAATCGTTTTAAAT 286

Qy 275 ccataatcaaaattttacatttttttaatacaaattttaaaattactcttattat 334

Db 287 ATATCTAGTTTTTAAAGAAATAAATTTAATTTAGTTTATATAAATTTTATTAT 346

Qy 335 agtatttttttttttttttagtctaaattataaaa-----ttatcttgaattttttgaa 389

Db 347 ATAGTATTTTATTTTATAAAATTAATTTTAAAGGATTAGCTTTAAATAAATTTT 406

Qy 390 ttttttaatttagtactaaataacataaatttttttctattctcagaataattcatt 449

Db 407 AATTTTAAATAAATATTTTAAATTAATGTAGCTTAAATAATAGTTATCATTAATAATTG 466

Qy 450 ttttaaaaaaattgatttttttagtataaattgtttgtataaattgaattgattt 509
 Db 467 TTATAATTTATTTTAAATTTTATTATTATTATTGTTAAATTAATTTATTATAATAT 526
 Qy 510 gaaagggacacaaatgaaaaaatgcttagaaaaaattctgtattcatcagctattta 569
 Db 527 AAATTTTAAATGATAAATAATTTATAAATTTAGATAAAATTTAGTATATTATTATATAAA 586
 Qy 570 t 570
 Db 587 T 587

Search completed: November 13, 1999, 12:50:53

Job time: 3068 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:57:41 ; Search time 104.22 Seconds
(without alignments)
95.226 Million cell updates/sec

Title: US-08-913-430-2
Perfect score: 2143
Sequence: 1 MKKMLRKKFLYSAYATSL.....NTILSPVIVTRANDNPDA 419

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2143	100.0	419	1 W01037	Mycoplasma 46-48 k
2	2091	97.6	419	1 R21829	Sequence of surfac
3	1437	67.1	281	1 R06279	Swine enzootic pne
4	684	31.9	132	1 R06277	Swine enzootic pne
5	631	29.4	120	1 R06278	Swine enzootic pne
6	138	6.4	29	1 W01034	Mycoplasma 46-48 k
7	122	5.7	277	1 Y00024	Enterococcus faeca
8	119.5	5.6	661	1 R42450	Enzyme involved in
9	119.5	5.6	661	1 R99458	Biosynthetic enzym
10	119	5.6	25	1 W01033	Mycoplasma 46-48 k
11	118	5.5	643	1 R70152	Streptococcus pneu
12	115.5	5.4	307	1 W80703	S. pneumoniae prot
13	113.5	5.3	2353	1 R99393	Haemophilus adhesi
14	112.5	5.2	3135	1 R57474	P. falciparum tran
15	109.5	5.1	994	1 R87018	Receptor tyrosine
16	109.5	5.1	994	1 W26366	Mouse Nuk tyrosine
17	109	5.1	594	1 R10005	Streptococcus CX78
18	108.5	5.1	323	1 R21083	GBP-delta-alpha1M
19	108.5	5.1	366	1 R21085	GBP-delta-alpha1M
20	107	5.0	859	1 R53268	120-128 kilodalton
21	107	5.0	1181	1 R53269	Tag A antigen of H
22	107	5.0	1181	1 R72593	H. pylori tagA ant
23	107	5.0	859	1 R72594	H. pylori tagA ant
24	107	5.0	722	1 R75705	Eph-related PTK Ce
25	107	5.0	710	1 R75714	Eph-related PTK Ce
26	107	5.0	1181	1 R91307	Helicobacter pylor
27	107	5.0	850	1 R91308	Helicobacter pylor
28	106	4.9	1098	1 R99392	Haemophilus adhesi
29	105	4.9	789	1 W46866	Bacillus thuringie
30	104	4.9	789	1 W46861	Bacillus thuringie
31	103.5	4.8	593	1 R07014	Protein G variant
32	103.5	4.8	789	1 W46869	Bacillus thuringie
33	103	4.8	776	1 W55105	Streptococcus pneu
34	103	4.8	876	1 Y04883	Mycobacterium spec
35	103	4.8	886	1 Y04884	Mycobacterium spec
36	102.5	4.8	991	1 R85090	EPH-like receptor
37	102	4.8	625	1 R14532	Protein C from mod
38	102	4.8	624	1 R14531	Protein C from pUC
39	102	4.8	789	1 W60216	Bacillus thuringie
40	101.5	4.7	789	1 W80320	Bacillus thuringie
41	100.5	4.7	809	1 R91261	VIP3A(a) synthetic
42	100.5	4.7	789	1 R91243	B. thuringiensis v
43	100.5	4.7	809	1 W19521	B. cereus VIP3A(a)

ALIGNMENTS

44	100.5	4.7	789	1 W19514	B. cereus VIP3A(a)
45	100.5	4.7	789	1 W46725	Vegetative insecti
RESULT	1				
ID	W01037	standard; Protein; 419 AA.			
AC	W01037				
DE	19-JAN-1997	(first entry)			
DT	Mycoplasma 46-48 kDa protective antigen.				
KW	Antigen; vaccine; mycoplasmal pneumonia; swine enzootic pneumonia;				
KW	diagnosis; antibody.				
OS	Mycoplasma hyopneumoniae strain Beaufort.				
FH	Key Location/Qualifiers				
FT	misc_difference 70	/note= "codon 70 in the nucleotide sequence is			
FT		a stop codon"			
FT	misc_difference 101	/note= "codon 101 in the nucleotide sequence is			
FT		a stop codon"			
FT	misc_difference 254	/note= "codon 254 in the nucleotide sequence is			
FT		is a stop codon"			
FT	W0628472-A1.				
PD	19-SEP-1996				
PF	15-MAR-1996; AU0149.				
PR	16-MAR-1995; AU-001789.				
PI	(UYME) UNIV MELBOURNE.				
DR	Doughty SW, Lee R, Walker J;				
DR	WPI: 96-433763/43.				
DR	N-PSDB; T38241.				
PT	Putative protective antigens against Mycoplasma - used for the				
PT	detection, prevention or treatment of Mycoplasma infections, esp. M.				
PT	hyopneumoniae in swine				
PS	Disclosure; Fig 7; 43pp; English.				
CC	The gene (T38241) coding for a 48 kDa putative protective antigen				
CC	(W01037) was obtd. from a Mycoplasma hyopneumoniae genomic library				
CC	by screening with a probe generated by PCR amplification (see also				
CC	T13814-16). The antigen had originally been isolated from M.				
CC	hyopneumoniae cells using antibody probes enriched with Mycoplasma-				
CC	specific antibodies. Other protective antigens were also identified				
CC	(see also W01024-31). Protective antigens and antibodies can be				
CC	used in vaccines for preventing or treating mycoplasmal infections,				
CC	Partic. M. hyopneumoniae infections in swine. They can also be used				
CC	for diagnosis.				
SQ	Sequence 419 AA;				
Query Match	100.0%;	Score 2143;	DB 1:	Length 419;	
Best Local Similarity	100.0%;	Fred. No. 1.3e-159;			
Matches 419;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MKKMLRKKFLYSAYATSLAIFAAAGCGGTSGTSDSKPQATLKHKYSNDISRI	60		
DB	1	MKKMLRKKFLYSAYATSLAIFAAAGCGGTSGTSDSKPQATLKHKYSNDISRI	60		
QY	61	ALTDPDNPRWISAQKDIISYVDETEAATSTITKNQDAQNMLTQQANLSPAPKGFIIAPE	120		
DB	61	ALTDPDNPRWISAQKDIISYVDETEAATSTITKNQDAQNMLTQQANLSPAPKGFIIAPE	120		
QY	121	NGSGVGTAVNTIADKGIPIVADRLITGSDKYDWMYVDFNEKVGELQGLSLAAGLLGKED	180		
DB	121	NGSGVGTAVNTIADKGIPIVADRLITGSDKYDWMYVDFNEKVGELQGLSLAAGLLGKED	180		
QY	181	GAFDSIDQWNEYLKSHPQETISFYTIAGSDNNQSYFYNGAMKYLKELMKNQNKIID	240		
DB	181	GAFDSIDQWNEYLKSHPQETISFYTIAGSDNNQSYFYNGAMKYLKELMKNQNKIID	240		
QY	241	LSPEGENAVYVPGWNYGTAGIQISFLTINKDPAGNKIKAVGSKPASIFKGLFAPNDGM	300		

Db 241 LSPGNAVYVPGWNYGTAGQRIQSFLTINKDPAGGNKIRAVGSKPASIFKGFAPNDGM 300
 Qy 301 AEQAITKLKLEGGFTQKIFVTRQDYNKAKTFIKDQDNMTIYKPDVKVGVAVEVLRVL 360
 Db 301 AEQAITKLKLEGGFTQKIFVTRQDYNKAKTFIKDQDNMTIYKPDVKVGVAVEVLRVL 360
 Qy 361 IAKNKASRSEVENELKAKLPNISFKYDQNTYKVGKNINTILVSPVITKANYDNPD 419
 Db 361 IAKNKASRSEVENELKAKLPNISFKYDQNTYKVGKNINTILVSPVITKANYDNPD 419

RESULT 2
 R21829
 ID R21829 standard; Protein; 419 AA.
 AC R21829;
 DT 02-NOV-1992 (first entry)
 DE Sequence of surface antigen 46kd.
 KW Swine pneumonia; epidemic; diagnosis; therapy.
 OS Mycoplasma hyopneumoniae.
 PN EP-475185-A.
 PD 18-MAR-1992.
 PF 27-AUG-1991; 114335.
 PR 27-AUG-1990; JP-224945.
 PA (NIFL-) NIPPON FLOUR MILLS.
 PI Seto Y, Futo S, Mitsuse S, Matsuo K, Tsuna M;
 DR WPI; 92-089874/12.
 DR N-PSDB; Q22042.
 PT DNA and peptide of mycoplasma hyopneumoniae - useful for
 PT diagnosis and treatment of swine mycoplasmal pneumonia
 PS disclosure; Page 4-6 and pages 34-36; 45pp; English.
 CC The inventors claim DNA encoding a surface antigen and primers used
 CC in a method of diagnosing mycoplasmal pneumonia of swine.
 CC Mycoplasmal cells were collected from the culture of M.hp, lysed in a
 CC buffer containing SDS, followed by the extraction of DNA and
 CC purification thereof. The DNA of M.hp is cleaved with a restriction
 CC enzyme HindIII, the resulting fragment is inserted into a plasmid
 CC pUC119 at the site cleaved with HindIII. This plasmid is transformed
 CC into E. coli as a host cell, colony hybridization is performed by the
 CC DNA probe method utilizing EcoRI fragments of pKUM1 and pKUM2 to
 CC select the bacterial cells containing a plasmid (pURR126) carrying the
 CC 46 kd antigenic gene. E. coli JM 109 which comprises the plasmid
 CC pKUM1 or pKUM2 is deposited with FRI under accession No. FERM P-10318
 CC or P-10319. The base sequence of the 46 kd antigen gene is given in
 CC Q22042.
 SQ Sequence 419 AA;

Query Match 97.6%; Score 2091; DB 1; Length 419;
 Best Local Similarity 98.3%; Pred. No. 1.5e-155;
 Matches 412; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 MKMLRKFLYSSAIYATSLASIIAFVAAGCGQTESGSTSDSKPQATLKHYSNDSIRI 60
 Db 1 MKMLRKFLYSSAIYATSLASIIAFVAAGCGQTESGSTSDSKPQATLKHYSNDSIRI 60
 Qy 61 ALTPDPNPRWISAQKDIISYDETEATSTITKNQDQNNWLTQOANLSPAPGFIITAE 120
 Db 61 ALTPDPNPRWISAQKDIISYDETEATSTITKNQDQNNWLTQOANLSPAPGFIITAE 120
 Qy 121 NGSGVGTAVNTIADKGIPIVAYDRLITGSKDYKDYVVSFDNEKVGELQGLSLAAGLLGKED 180
 Db 121 NGSGVGTAVNTIADKGIPIVAYDRLITGSKDYKDYVVSFDNEKVGELQGLSLAAGLLGKED 180
 Qy 181 GAFDSIDQMEYLSHNPQETISFYTTAGSODDNNNSQYFYNGAMKVLKELMKNQKIID 240
 Db 181 GAFDSIDQMEYLSHNPQETISFYTTAGSODDNNNSQYFYNGAMKVLKELMKNQKIID 240
 Qy 241 LSPGNAVYVPGWNYGTAGQRIQSFLTINKDPAGGNKIRAVGSKPASIFKGFAPNDGM 300
 Db 241 LSPGNAVYVPGWNYGTAGQRIQSFLTINKDPAGGNKIRAVGSKPASIFKGFAPNDGM 300
 Qy 301 AEQAITKLKLEGGFTQKIFVTRQDYNKAKTFIKDQDNMTIYKPDVKVGVAVEVLRVL 360
 Db 301 AEQAITKLKLEGGFTQKIFVTRQDYNKAKTFIKDQDNMTIYKPDVKVGVAVEVLRVL 360

Db 301 AEQAITKLKLEGGFTQKIFVTRQDYNKAKTFIKDQDNMTIYKPDVKVGVAVEVLRVL 360
 Qy 361 IAKNKASRSEVENELKAKLPNISFKYDQNTYKVGKNINTILVSPVITKANYDNPD 419
 Db 361 IAKNKASRSEVENELKAKLPNISFKYDQNTYKVGKNINTILVSPVITKANYDNPD 419

RESULT 3
 R06279
 ID R06279 standard; protein; 281 AA.
 AC R06279;
 DT 19-DEC-1990 (first entry)
 DE Swine enzootic pneumonia mycoplasma surface antigen polypeptide.
 KW Mycoplasmal pneumonia; enzootic pneumonia; ds.
 OS Mycoplasma hyopneumoniae.
 PN J02167079-A.
 PD 27-JUN-1990.
 PF 21-DEC-1988; 322829.
 PR 21-DEC-1988; JP-322829.
 PA (NISE-) NIPPON SEIFUN KK.
 DR WPI; 90-241949/32.
 DR N-PSDB; Q05578.
 PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
 PT for early detection and treatment of mycoplasmal or enzootic
 PT pneumonia of pigs
 PS Claim 1; Page 570; 28pp; Japanese.
 CC Sequence encoding polypeptides may be used as probes for the early
 CC diagnosis of pneumonia in pigs. The sequences may be used to produce
 CC an expression vector and transform a suitable host.
 SQ Sequence 281 AA;

Query Match 67.1%; Score 1437; DB 1; Length 281;
 Best Local Similarity 99.6%; Pred. No. 9.3e-105;
 Matches 280; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 139 IVAYDRLITGSDKYVVSFDNEKVGELQGLSLAAGLLGKEDGAFDSIDQMEYLSHNP 198
 Db 1 IVAYDRLITGSDKYVVSFDNEKVGELQGLSLAAGLLGKEDGAFDSIDQMEYLSHNP 60
 Qy 199 QETISYTTAGSODDNNNSQYFYNGAMKVLKELMKNQKIIDLSPEGENAVYVPGWNYGT 258
 Db 61 QETISYTTAGSODDNNNSQYFYNGAMKVLKELMKNQKIIDLSPEGENAVYVPGWNYGT 120
 Qy 259 AGORIOSFLTINKDPAGGNKIRAVGSKPASIFKGFAPNDGMAEQAITKLKLEGGFTQKI 318
 Db 121 AGORIOSFLTINKDPAGGNKIRAVGSKPASIFKGFAPNDGMAEQAITKLKLEGGFTQKI 180
 Qy 319 FVTRQDYNKAKTFIKDQDNMTIYKPDVKVGVAVEVLRVLIATKNKASRSEVENELKA 378
 Db 181 FVTPQDYNKAKTFIKDQDNMTIYKPDVKVGVAVEVLRVLIATKNKASRSEVENELKA 240
 Qy 379 KLPNISFKYDQNTYKVGKNINTILVSPVITKANYDNPD 419
 Db 241 KLPNISFKYDQNTYKVGKNINTILVSPVITKANYDNPD 281

RESULT 4
 R06277
 ID R06277 standard; protein; 132 AA.
 AC R06277;
 DT 19-DEC-1990 (first entry)
 DE Swine enzootic pneumonia mycoplasma surface antigen polypeptide.
 KW Mycoplasmal pneumonia; enzootic pneumonia; ds.
 OS Mycoplasma hyopneumoniae.
 PN J02167079-A.
 PD 27-JUN-1990.
 PF 21-DEC-1988; 322829.
 PR 21-DEC-1988; JP-322829.
 PA (NISE-) NIPPON SEIFUN KK.
 DR WPI; 90-241949/32.
 DR N-PSDB; Q05576.
 PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used

PT for early detection and treatment of mycoplasma or enzootic
 PT pneumonia of pigs
 PS Claim 1; Page 570; 28pp; Japanese.
 CC Sequence encoding polypeptides may be used as probes for the early
 CC diagnosis of pneumonia in pigs. The sequences may be used to produce
 CC an expression vector and transform a suitable host.
 SQ Sequence 132 AA;

Query Match 31.9%; Score 684; DB 1; Length 132;
 Best Local Similarity 100.0%; Pred. No. 1.8e-46;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 159 DNEKVELQGLSLAAGLLGKEDGAFDSIDQNEYLYKSHMPQETISFYTIAGSODDNNQY 218
 Db 1 DNEKVELQGLSLAAGLLGKEDGAFDSIDQNEYLYKSHMPQETISFYTIAGSODDNNQY 60
 QY 219 FYGAMKVLKELMKNQNKIIDLSPEGENAVYVPGWNYGTAGORIQSFLTIKNDPAGGK 278
 Db 61 FYGAMKVLKELMKNQNKIIDLSPEGENAVYVPGWNYGTAGORIQSFLTIKNDPAGGK 120
 QY 279 IKAVGSKPASIF 290
 Db 121 IKAVGSKPASIF 132

RESULT 5
 R06278 ID R06278 standard; protein; 120 AA.
 AC R06278;
 DT 19-DEC-1990 (first entry)
 DE Swine enzootic pneumonia mycoplasma surface antigen polypeptide.
 KW Mycoplasma pneumoniae; enzootic pneumonia; ds.
 OS Mycoplasma hyopneumoniae.
 PN J02167079-A.
 PD 27-JUN-1990.
 PF 21-DEC-1988; 322829.
 PR 21-DEC-1988; JP-322829.
 PA (NISE-) NIPPON SEIFUN KK.
 DR WPI: 90-241949/32.
 DR N-PSDB: Q05577.
 PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
 PT for early detection and treatment of mycoplasma or enzootic
 PT pneumonia of pigs
 PS Claim 1; Page 570; 28pp; Japanese.
 CC Sequence encoding polypeptides may be used as probes for the early
 CC diagnosis of pneumonia in pigs. The sequences may be used to produce
 CC an expression vector and transform a suitable host.
 SQ Sequence 120 AA;

Query Match 29.4%; Score 631; DB 1; Length 120;
 Best Local Similarity 100.0%; Pred. No. 2.2e-42;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 153 DWYVSFDEKVGELQGLSLAAGLLGKEDGAFDSIDQNEYLYKSHMPQETISFYTIAGSQD 212
 Db 1 DWYVSFDEKVGELQGLSLAAGLLGKEDGAFDSIDQNEYLYKSHMPQETISFYTIAGSQD 60
 QY 213 DNNQSYFYNGAMKVLKELMKNQNKIIDLSPEGENAVYVPGWNYGTAGORIQSFLTIKND 272
 Db 61 DNNQSYFYNGAMKVLKELMKNQNKIIDLSPEGENAVYVPGWNYGTAGORIQSFLTIKND 120
 RESULT 6
 W01034 ID W01034 standard; Peptide; 29 AA.
 AC W01034;
 DT 19-JAN-1997 (first entry)
 DE Mycoplasma 46-48 kDa protective antigen fragment CNBR F1.
 KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumonia;
 KW diagnosis; antibody.
 OS Mycoplasma hyopneumoniae strain Beaufort.

PN W09628472-A1.
 PD 19-SEP-1996.
 PF 15-MAR-1996; AU0149.
 PR 16-MAR-1995; AU-001789.
 PA (UYME) UNIV MELBOURNE.
 PI Doughty SW, Lee R, Walker J;
 DR WPI: 96-433763/43.
 PT Putative protective antigens against Mycoplasma - used for the
 PT detection, prevention or treatment of Mycoplasma infections, esp. M.
 PT hyopneumoniae in swine
 PS Claim 13; Page 28; 43pp; English.
 CC A 46-48 kDa putative protective antigen against Mycoplasma contains
 CC the N-terminal sequence given in W01034 and the internal CNBR
 CC fragments given in W01034-36. The antigen was isolated from
 CC Mycoplasma hyopneumoniae cells using antibody probes enriched with
 CC Mycoplasma-specific antibodies. Other protective antigens were
 CC also identified (see also W01024-32). A gene (T38241), coding for
 CC the 48 kDa antigen (W01037) was isolated from a genomic library.
 CC Protective antigens and antibodies can be used in vaccines for
 CC preventing or treating mycoplasma infections, partic. M.
 CC hyopneumoniae infections in swine. They can also be used for
 CC diagnosis.
 SQ Sequence 29 AA;

Query Match 6.4%; Score 138; DB 1; Length 29;
 Best Local Similarity 100.0%; Pred. No. 8.8e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 TTYKPKVLGKVAVEVLRLVIAKKNKASR 369
 Db 1 TTYKPKVLGKVAVEVLRLVIAKKNKASR 29

RESULT 7
 Y00024 ID Y00024 standard; Protein; 277 AA.
 AC Y00024;
 DT 20-APR-1999 (first entry)
 DE Enterococcus faecalis protein EF009.
 KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KW detection; attenuation; antigenic.
 OS Enterococcus faecalis.
 PN W09850554-A2.
 PD 12-NOV-1998.
 PF 04-MAY-1998; U08959.
 PR 14-NOV-1997; US-066009.
 PR 06-MAY-1997; US-044031.
 PR 16-MAY-1997; US-046655.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
 DR WPI: 99-070095/06.
 DR N-PSDB; X20014.
 PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection
 PS Claim 9; Page 86; 301pp; English.
 CC The present sequence represents a protein isolated from
 CC Enterococcus faecalis. The present invention describes genes, proteins
 CC and antigenic polypeptides isolated from E. faecalis. The proteins can
 CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.
 SQ Sequence 277 AA;

Query Match 5.7%; Score 122; DB 1; Length 277;
 Best Local Similarity 22.8%; Pred. No. 0.037;
 Matches 71; Conservative 43; Mismatches 107; Indels 90; Gaps 16;

Query Match	5.6%	Score 119.5	DB 1	Length 661
Best Local Similarity	18.3%	Pred. No. 0.2		
Matches	97	Conservative	73	Mismatches 206
				Indels 153
				Gaps
QY	15	IYATSLASIIAFVAAAGCGTESGTSDSKPOAFTLKHVS-NDSIRIALTDPDNPWISA	73	
Db	5	LMAISIMLSFVNALAAQHEHDHITVDYEGKAT-EHTAHNQAVAKTLNFADTRAPEOS	63	
QY	74	OKDIISVXDTEAATSTITKNQ-----DAONNLWTOQANLSPAPKGF-----IAP	119	
Db	64	SKNLVAKFDK--ATADILRAEFASIDEIPDSVNPNSLYRQALNMPVNGLYKVSQDIQ	120	
QY	120	ENGSGVGTAVNTADKGIPVAVDRLIT-GSDKYDWTVSPDN-EKVGEQLGSLA----	173	
Db	121	VRGTDLSNLTIRSDNG--WIAVDVLTTRKAAKASIQFALKNLPKQGDLPVMAIYSHSH	178	
QY	173	AGLKGEDGAFDSIDQMNEYLKSHMPQETISFYTTIAGQQDDNNSOFYNGAM-----	225	

[illegible]

[illegible]

RESULT	10	
W01033		
AC	W01033 standard; Peptide; 25 AA.	
AD	W01033;	
AT	19-JAN-1997 (first entry)	
DE	Mycoplasma 46-48 kDa protective antigen N-terminal peptide.	
KW	Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumoniae;	
KW	diagnosis; antibody.	
OS	Mycoplasma hyopneumoniae strain Beaufort.	
FT	Key	Location/Qualifiers
FT	misc_difference 3	
FT	/note= "undetermined amino acid"	
PN	W09628472-AL.	
PD	19-SEP-1996.	
PF	15-MAR-1996; AU0149.	
PR	16-MAR-1995; AU-001789.	
PA	(UYME) UNIV MELBOURNE.	
PI	Doughty SW, Lee R, Walker J;	
PI	WPI; 96-433763/43.	
PT	Putative protective antigens against Mycoplasma - used for the	
PT	detection, prevention or treatment of Mycoplasma infections, esp. M.	
PT	hyopneumoniae in swine	
PT	Claim 12; Page 28; 43pp: English.	
PS	A 46-48 kDa putative protective antigen against Mycoplasma contains	
CC	the N-terminal sequence given in W01033 and the internal CNBR	
CC	fragments given in W01034-36. The antigen was isolated from	
CC	Mycoplasma hyopneumoniae cells using antibody probes enriched with	
CC	Mycoplasma-specific antibodies. Other protective antigens were	
CC	also identified (see also W01024-32). A gene (T38241) coding for	
CC	the 48 kDa antigen (W01037) was isolated from a genomic library.	
CC	Protective antigens and antibodies can be used in vaccines for	
CC	preventing or treating mycoplasma infections, partic. M.	
CC	hyopneumoniae infections in swine. They can also be used for	
CC	diagnosis.	
CC	Sequence 25 AA;	
CC	Sequence 25 AA;	

Query Match	5.6%	Score 119;	DB 1;	Length 25;
Best Local Similarity	96.0%;	Pred. No. 0.0022;		
Matches 24;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

```

QY      29 AGCGQTESGSTDSPKQAETLKHKV  53
        |||||||
Db       1 AGXGQTESGSTSDSPKQAETLKHKV  25

RESULT  11
R70152
ID      R70152 standard; Protein; 643 AA.
AC      R70152;
DT      14-FEB-1996 (first entry)

```

Streptococcus pneumoniae strain SPR098 pIpA. pIpA; export protein; pIpi; AmiA; virulence determinant; permease like protein; penicillin binding protein 1A; pyruvate oxidase; regulatory element; acellular vaccine; antibody.

Streptococcus pneumoniae.

WO9506732-A2.

09-MAR-1995.

01-SEP-1994; U099942.

01-SEP-1993; US-116541.

18-MAY-1994; US-245511.

(UYRO) UNIV ROCKEFELLER.

Masure HR, Pearce BJ, Tuomanen E; WPI; 95-115448/15.

N-PSDB; Q83240.

Novel gene fragments encoding specific bacterial exported proteins - specifically of S. pneumoniae, useful as vaccines

Claim 12; Page 116-8; 169pp; English.

This sequence represents pIpA (permease like protein A). This sequence is also referred to as pIpi. This is an exported protein of S.pneumoniae. Export proteins are the proteins in pathogenic bacteria that are virulence determinants. The sequences encoding pIpA and the downstream region for this gene were obtained by chromosome walking. The upstream region of the coding sequence was obtained by homology cloning using the sequences shown in Q82533-4. pIpA is not thought to be part of an operon and encodes a putative substrate binding protein responsible for peptide transport. The pIpA protein shows 80% similarity to AmiA (another substrate binding protein). pIpA is thought to serve as a regulatory element that modulates the cell density dependent induction of competence by mediating the binding and transport of signalling molecules. The DNA encoding this sequence can be inserted into an expression vector (preferably a bacterial expression vector) to provide for high levels of expression of the protein. The protein can then be used in the production of an acellular vaccine. These vaccines are used to provide protection from Gram positive bacterial infection. Antibodies against export proteins can be used for diagnosis of infection and in passive immune therapy.

Sequence 643 AA;

```

Query Match      5.5%; Score 118; DB 1: Length 643;
Best Local Similarity 21.1%; Pred. No. 0.25;
Matches 90; Conservative 60; Mismatches 163; Indels 114; Gaps 20;

Qy      27 VAACGGTSGSTSDSKPOAETLKHKYSNDSIRALTDPDNPWISACKDILISVDETEA 86
Db      2 VLAAC-----SGSGSAGE-KTFSY-----IYETDPDN-----LNTLTAKA 38

Qy      87 ATSTITRN-----ODAQNWLTOQANLSPAPKGFIIAPENGSGVGTAVNTIADKGIPI 139
Db      39 ATANITSNVYDGLLENDRYGNFVPSMA-----EDWS-----VSKDGL-- 76

Qy      140 VAYDRLTGSDKYDWYVSFDNEKVEGLOGLSLAAGLICKEDGAFDSDOMNEYLK---SH 196
Db      76 -TYTYTIRKDAK--WYTS-EGEYYAAVKAQDFVTGLKYAADKKSDALYPVOESIKGLDAY 131

Qy      197 MPQETISFYTIA-GSODNNNSQFYNGAMKVLKELMKNSONKIIDLSPEGENAVVVGWN 255
Db      132 VKGEEKFSQVGIKALDEQTVQYTLN-----KPSFWNSKTTMGVLAPVNEEFLNKSGDD 186

Qy      256 YGTAGORIQSFLLTINKDPA-----GGNKIRAVGSKPASIFKGF LAPNDGMAEQ---AATK 307
Db      187 FAKA-----TDPSSLLYNGPYLLKSIVTKSSVEF---AKPNPYWKDKNVHIDK 231

Qy      308 LKLGSDFTQKIFVTRQDYNKAKT-----FIKGDQNMITYKDPKVLGKVAV 354
Db      232 VKLSFWGDQDTSKPAENFKDGLSFAARLYPTSAFAELEKSMKDNIVYTOQDSITYLVGT 291

Qy      355 EVLAVLIAKKNKASRSEVENELKALPN-----TSFYXDQNTYKQV-----GKNINTI 402
Db      292 NIDRQSYKYTSKDSDEQKASTKKALLNKDFQATAFGFDRTAYASQLNGOTGASKILRN 351

Qy      403 LVSVPVIV 409
Db      109

```


Db 703 VMIITFENGSLDSEFLRQNDGQFTVIQ-----LVGMLRGIAAGMKYLADMNYVHRD 754
Qy 376 LKAK 379
| |
Db 755 LAAR 758

Search completed: November 13, 1999, 18:59:43
Job time: 122 sec

Result No.	Query			ID	Description	
	Score	Match	Length			
1	119.5	5.6	661	1	US-08-375-709-3	Sequence 3, Appli
2	119.5	5.6	661	2	US-08-753-929-3	Sequence 3, Appli
3	118	5.5	643	2	US-08-245-511-47	Sequence 47, Appl
4	112.5	5.2	3135	1	US-08-323-170B-2	Sequence 2, Appli
5	107	5.0	1859	1	US-08-053-614-2	Sequence 2, Appli
6	107	5.0	1181	1	US-08-053-614-4	Sequence 4, Appli
7	107	5.0	722	1	US-08-163-809-4	Sequence 4, Appli
8	107	5.0	710	1	US-08-162-809-22	Sequence 22, Appli
9	107	5.0	859	1	US-08-316-397B-2	Sequence 2, Appli
10	107	5.0	1181	1	US-08-316-397B-4	Sequence 4, Appli
11	107	5.0	859	2	US-09-034-306-2	Sequence 2, Appli
12	107	5.0	1181	2	US-09-034-306-4	Sequence 4, Appli
13	107	5.0	859	3	PCT-US93-09782-2	Sequence 4, Appli
14	107	5.0	1181	3	PCT-US93-09782-4	Sequence 4, Appli
15	106.5	5.0	1912	1	US-08-409-995-4	Sequence 4, Appli
16	106	4.9	1098	1	US-08-409-995-2	Sequence 2, Appli
17	102.5	4.8	991	3	PCT-US95-04681-13	Sequence 13, Appli
18	101	4.7	995	2	US-08-673-789-5	Sequence 5, Appli
19	100.5	4.7	789	2	US-08-471-033-29	Sequence 29, Appl
20	100.5	4.7	789	2	US-08-471-044-29	Sequence 29, Appl
21	100.5	4.7	789	2	US-08-463-403A-29	Sequence 29, Appl
22	100.5	4.7	789	2	US-08-471-046A-29	Sequence 29, Appl
23	100.5	4.7	789	2	US-08-470-566B-29	Sequence 29, Appl
24	100.5	4.7	879	2	US-08-838-219B-2	Sequence 2, Appli
25	99	4.6	877	2	US-08-673-789-2	Sequence 2, Appli
26	98.5	4.6	3289	1	US-08-477-451-2	Sequence 2, Appli
27	98	4.6	995	1	US-08-162-809-18	Sequence 18, Appl
28	97.5	4.5	970	3	PCT-US95-04681-11	Sequence 11, Appl
29	96	4.5	1702	3	US-08-838-219B-6	Sequence 6, Appli
30	96	4.5	746	2	PCT-US95-10661A-5	Sequence 5, Appli
31	95	4.4	744	1	US-08-162-809-20	Sequence 20, Appl
32	95	4.4	928	2	US-08-442-248-2	Sequence 2, Appli
33	95	4.4	928	2	US-08-440-815-2	Sequence 2, Appli
34	95	4.4	488	2	US-08-797-236-2	Sequence 2, Appli
35	94.5	4.4	689	1	US-08-248-021A-2	Sequence 2, Appli
36	94.5	4.4	970	1	US-08-373-709-7	Sequence 7, Appli
37	94.5	4.4	970	2	US-08-752-929-7	Sequence 7, Appli
38	94.5	4.4	908	2	US-08-249-380-2	Sequence 2, Appli
39	94.5	4.4	978	2	US-08-415-593-43	Sequence 43, Appl

```
Query Match          5.6%; Score 119.5; DB 1; Length 661;
Best Local Similarity 18.3%; Pred. No. 0.011;
Matches 97; Conservative 73; Mismatches 206; Indels 153; Gaps 20;

15 IYATSLASIIAFVACCGGTGSGTSDSPQAEITLKHYS-NDSIRIALTDPPNPRISA 73
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
5 LMAISIMLSFENALAAQHEDHITVDYEGKAAT-EHTIAHQAVAKTLNFADTRAFEQS 63
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
74 QRDIISSYDTEEAAYSTIKNQ-----DAQNNWLTCQAANLSPAPKGF I----IAP 119
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
```

Db 64 SKNLVAFEDK---ATADILRAEFATISDEIPDSVNPISLYRQALNMPNGLYKVSDEGIYQ 120
QY 120 ENGSGVGTAVNTIADKGIPIVAYDRLIT-GSKDYDWYSFDN-EKVGELOGLSLA----- 173
Db 121 VRGTDLSNLTIRSDNG--WYAYDVLTLKEAKASLOFALKNLPKDGLPVMYIYSHSH 178
QY 173 AGLLGKEDGAFDSIDOMNEYLKSHMPQETISFYTIAGSQDDNNQSYFYNGAM----- 225
Db 179 ADHFGGARGVQEMFDPVKVYSDNITKEIVDENVLGNAMSRRAAYQYGATLKGHDHGIV 238
QY 225 -----KVLKELMNKSNQK-----IIDLS---PEGNAVYVPG-- 254
Db 239 DAALGKGLSGEITYVAPDYTLNSEGKWEYTLIDGLEWFMFDASGTEASEMITYIPSKK 298
QY 254 --WN-----YGTAGORISFLTINKDPAGGNKIKAVGSKPASIFKGLAP----- 297
Db 299 ALWTAELTYOGMNIYTLRGKVRDALKWSKDI--NEMINAFQDVEVLFASHAPVWGN 356
QY 297 -----NDGMAEQAITKLEGFDTQKIFVTRQD-----YN 326
Db 357 QAINDFLRLQRDNYGLVHNOTLRLANDGVGIQDIDGAIQDTIPESYKWTHTNGYHGYS 416
QY 327 DKAKT-----FIKDGQDNWTYKPKVLGKVAVEV 356
Db 417 HNKAAYNKYLYGFDMPANLPLTKQESAKFVEYMGGADAIAKRAKDDYAOGE----- 472
QY 357 LRVLIANKKASRSEVENELKAPLNISFYDNTQYKVGKNINTILVS 405
Db 472 YRFVATALNKVMAEPENDSARQL--LADTYEQLGYQAGAGWRNIYLT 518

RESULT 2
US-08-752-929-3
; Sequence 3, Application US/08752929
; Patent No. 5798259
; GENERAL INFORMATION:
; APPLICANT: YAZAWA, Kazunaga
; APPLICANT: YAMADA, Akiko
; APPLICANT: KATO, Seishi
; APPLICANT: KONDO, Kiyosi
; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing
; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,929
; FILING DATE: 20-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/375,709
; FILING DATE: 20-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/178,251
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-147945
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 661 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-752-929-3

Query Match 5.6%; Score 119.5; DB 2; Length 661;
Best Local Similarity 18.3%; Pred. No. 0.011;
Matches 97; Conservative 73; Mismatches 206; Indels 153; Gaps 20;

QY 15 IYATSLASIIAFVAAGCGQTESGSDSKPQAEITLKHVS-NDISRIALTDPDNPWISA 73
Db 5 LMAISIMSLFSFNALAAQHEHDHITVDYEGKAAT-EHTIAHQAVAKTLNFAADTRAPEQS 63
QY 74 OKDIISYVDETERATSTITKNQ-----DAQNWLTOQANLSPAPKGF-----IAP 119
Db 64 SKNLVAFEDK---ATADILRAEFATISDEIPDSVNPISLYRQALNMPNGLYKVSDEGIYQ 120
QY 120 ENGSGVGTAVNTIADKGIPIVAYDRLIT-GSKDYDWYSFDN-EKVGELOGLSLA----- 173
Db 121 VRGTDLSNLTIRSDNG--WYAYDVLTLKEAKASLOFALKNLPKDGLPVMYIYSHSH 178
QY 173 AGLLGKEDGAFDSIDOMNEYLKSHMPQETISFYTIAGSQDDNNQSYFYNGAM----- 225
Db 179 ADHFGGARGVQEMFDPVKVYSDNITKEIVDENVLGNAMSRRAAYQYGATLKGHDHGIV 238
QY 225 -----KVLKELMNKSNQK-----IIDLS---PEGNAVYVPG-- 254
Db 239 DAALGKGLSGEITYVAPDYTLNSEGKWEYTLIDGLEWFMFDASGTEASEMITYIPSKK 298
QY 254 --WN-----YGTAGORISFLTINKDPAGGNKIKAVGSKPASIFKGLAP----- 297
Db 299 ALWTAELTYOGMNIYTLRGKVRDALKWSKDI--NEMINAFQDVEVLFASHAPVWGN 356
QY 297 -----NDGMAEQAITKLEGFDTQKIFVTRQD-----YN 326
Db 357 QAINDFLRLQRDNYGLVHNOTLRLANDGVGIQDIDGAIQDTIPESYKWTHTNGYHGYS 416
QY 327 DKAKT-----FIKDGQDNWTYKPKVLGKVAVEV 356
Db 417 HNKAAYNKYLYGFDMPANLPLTKQESAKFVEYMGGADAIAKRAKDDYAOGE----- 472
QY 357 LRVLIANKKASRSEVENELKAPLNISFYDNTQYKVGKNINTILVS 405
Db 472 YRFVATALNKVMAEPENDSARQL--LADTYEQLGYQAGAGWRNIYLT 518

RESULT 3
US-08-752-929-3
; Sequence 47, Application US/08245511
; Patent No. 5928900
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:

[illegible]

Query Match	5.0%	Score 107;	DB 1;	Length 1181;
Best Local Similarity	22.0%	Pred. No. 0.38;		
Matches	93;	Conservative	55;	Mismatches 137; Indels 138; Gaps 20;
Qy	41	DSKPOAETLKH-KVSNDSIRIALTDPDNPRIWISAQKDIIISGVDETEATSTIITKNDAQN	99	
Db	152	DDKEKAEFLSAKOSFAGIIIG-----NOIRFDQKMGVFDE-----SLKERQSAEK	198	
Qy	100	N-----WLTQOANLSPAPKGFIIAPENGSGVGTAVN-----TIADKGP	138	
Db	199	NGGPTGGDWL--DIFLS-----FIFDKKQSSDYKEAINQEPLPHVQPDIASTTHIQGLP	251	
Qy	139	IVAYDRLIITGSDKYDWTVSDNEKVGELGI-----SLAGLL	176	
Db	252	PESRDLL-----DERGNFSKFTLGDMEDLDVEGVADMDPNKYFNQILLIHNNLTSSVLM	304	
Qy	177	GKEDGAFDSIDQNEYLKSHMPOETISFYITIA-----GSODDNNSSQYFV-----NGAMKVL	227	
Db	305	GSHDG-----IEPKVSLLYAGNGGGFGAKHDDWNAVTVYKQOQGNVATII	349	
Qy	228	KELMKNQNKIIDLSPEGEN--AVYVPGWNYGTAGQRIQSFLTINKDPAGNKKIAGVSK	285	

```

Query Match      5.0%; Score 107; DB 1; Length 859;
Best Local Similarity 22.0%; Pred. No. 0.23;
Matches 93; Conservative 55; Mismatches 137; Indels 138; Gaps 20;

QY 41 DSKPQATLKH-KVNSDSIRIALDPDPNPRWISAKDIIISVDETEATSTITNKDAQN 99
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 DDKEAKEFLSAKQSAGIILG-----NOIRTDQKPMGVFDE-----SLKREQAEK 198

QY 100 N-----WLTQOANLSPAPGFIIAPENGSGVGTAVN-----TADKGIP 138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 NGPGTGGDWL--DIFLS-----FIDFKQSSDVKAEAINQEPLPHVQPDIASTTHIQGL 251

QY 139 IVAYDRLLITGSDKYDVYSPFNEKVGELQGL-----SLAAGLL 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 PESRDLL-----DERGNFSKFTLGDMEMLDVEGVADMDPNFYKNQLLIHNLTSSVLM 304

QY 177 GKEDGAFDSIDOMNEYLKSHMPQETISYTTA---GSQDDNNSQYFY----NGAMKVL 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 GSHDP-----IEPEKVSLLYAGNGGFGAKHDWNAITVGYKDQOGNNVATII 349

QY 228 KELMKNQNKIIDLSPEGEN--AVVPQWNYGTAGORIQSFLLTINKDPAGGNKIKAVGSK 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 350 NVHMKNGSLVIAGEGKGINNPSTLYKEDQLTUSQRALSQEEIQ-----NKID----- 399

QY 286 PASIFKGFAPLN-----DGMARQAITKLKLEGFDTQK-----IFVTRQDYND 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 ----FMFEALQNNAKLDSLSSEKEKFPNEIKDFQKDSKPYLDALGNDRIAFVSKDPPH 454

QY 328 KAK-TFIKDGDQNNMTIYKPKDVLGVKVAVEVLRLVLIANKKNASRSBEVELKAKLPNISFK 386

```

Db 350 NVHMKNGSLVIAGGEGKINNPSFYLYKEDQLTGSRALSQBEIQ-----NKID----- 399

QY 286 PASIFKGLAPN---DGMAEQAITKLKLEGEFTQK-----IFVTRODYND 327

Db 399 -----FMEFLAQNNAKLDSLSEKEKEFRNEIKDFQKDSRPLYDALGNDRIAFVSKDKPKH 454

QY 328 KAK-TFIKGDQNMITYKPKVLGVKAVVEVLRLVLIAKKNKASRSEVENELKAKLPNISFK 386

Db 455 SALITEFNRKGDLSYTL---KVMGK---KQIKALDRKKNVTLQGNLKHGDMFVFNYSFK 507

QY 387 YDN 389

Db 508 YDN 510

RESULT 7

US-08-162-809-4

; Sequence 4, Application US/08162809

; Patent No. 5457048

; GENERAL INFORMATION:

; APPLICANT: Pasquale, Elena B.

; APPLICANT: Sajjadi, Fereydoun G.

; TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES, AND METHODS OF USE

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CAMPBELL AND FLORES

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States of America

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/162,809

; FILING DATE:

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-LJ 9503

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 722 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-162-809-4

Query Match 5.0%; Score 107; DB 1; Length 722;

Best Local Similarity 19.3%; Pred. No. 0.18;

Matches 83; Conservative 59; Mismatches 168; Indels 120; Gaps 18;

QY 32 GQTESGSDS-KPOAETLKHVNSDRIALTDPDNPRWISAQKDIISYVDTEAATST 90

Db 165 GYEENNTQAPSPVSSVKKKITSLSLWQEPDRPNGIILEYE-IKFEKQDQSTST 223

QY 91 IYKQDAQNWLTOQANLSPAPKGFIIAPENGSGVGTAVNTADKGIPIVADRLITGSD 150

Db 224 IIKSKETA---ITADGLKPGSAIVFOIRARTAGYGGFRFEFTSPVLA-----ASSD 275

QY 151 KYDWYVSFNEKVGELQGLSLAAGLL--GKEDGAFDSIDQMNEYLKSHMPQETISFYTTA 208

Db 276 QSOIPIIVSVTVGVIL-LAVVIGFLLSRRCG-----YSKA 311

QY 209 GSQDDNNQSYFYNGAMKVLKELMKNQNKIIDLSPEGENAVYVPGWNYGTAGRIQSF-- 267

Db 312 KQDPEEKMHFNHGHKL-----PGVR-TYIDPHTYEDPNQAVHEFAK 353

QY 267 -----LTINKDPAGS-----NLIKAVGSK-----PASIF 290

Db 354 ETEASCITIERVIGAGEFGEVCSGRKLQGRKREFVPAKTLKVGYTEKORRDFLGEASIM 413

QY 291 KGFAPNDGMAQAIATKLKLEGEFTQKIFVTRODYNDKAKTFIKGDQNMITYKPKVLG 350

Db 414 GQFDHPNIIHLEGGVTKSK-----PVMIVTEYMENGSIDTFLKKNQDGOFTVIQ----- 462

QY 351 KVAVEVLR-VLIAKNKASRSEVENELKAKLPNISFKYVDNMITYKVGKNINTILVSPVIV 409

Db 462 --LVGMLRGIASGMKYLSDMGVYHRDLAAR--NILIN-SNLVCKVSDFGLSRVL----- 511

QY 410 TKANVDNPD 419

Db 511 -----EDDPEA 516

RESULT 8

US-08-162-809-22

; Sequence 22, Application US/08162809

; Patent No. 5457048

; GENERAL INFORMATION:

; APPLICANT: Pasquale, Elena B.

; APPLICANT: Sajjadi, Fereydoun G.

; TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES, AND METHODS OF USE

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CAMPBELL AND FLORES

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States of America

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/162,809

; FILING DATE:

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-LJ 9503

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 710 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-162-809-22

Query Match 5.0%; Score 107; DB 1; Length 710;

Best Local Similarity 19.3%; Pred. No. 0.17; 168; Indels 120; Gaps 18;

Matches 83; Conservative 59; Mismatches 168; Indels 120; Gaps 18;

QY 32 GQTESGSDS-KPOAETLKHVNSDRIALTDPDNPRWISAQKDIISYVDTEAATST 90

Db 165 GYEENNTQAPSPVSSVKKKITSLSLWQEPDRPNGIILEYE-IKFEKQDQSTST 223

QY 91 IYKQDAQNWLTOQANLSPAPKGFIIAPENGSGVGTAVNTADKGIPIVADRLITGSD 150


```
Db 224 IIKSKETA---ITADGLKPGSAYVQIRARTAAAGYGGFSRRFFETSPVLA-----ASSD 275
QY 151 KYDWYVDFNEKVEGLOGLSLAAGLL--GKEDGAFDSIDQWNEVYLKSHMPOETISFYTIA 208
Db 276 QSQPIIIVSVTVGVIL-LAVVIGFLLSGRCG-----YSKA 311
QY 209 GSODDNNQSYFYNGAMKVLKELMKNSQNKIIDLSPEGENAVYVPGWNYGTAGORISQF-- 267
Db 312 KQDPEEEKMHFHNGHKL-----PCVR-TYIDPHYEDPNQAVHEFAK 353
QY 267 -----LTINKDPAGG-----NKIKAVGSK-----PASIF 290
Db 354 EIEASCITIERVIGAGVEVSGRLKQKREFPVAIKLVGYTEKQRDRFLGEASIM 413
QY 291 KGFLAPDGMAGAEQAITKLEGEFTQRIEFTVTRDYNKAKTFIKDQDNMTIYKPKDKVLG 350
Db 414 QFDPHNIHLEGVYTKS-----PWIIVTEYENGSLDTFLKNDGQTVTIQ----- 462
QY 351 KVAVEVLR-VLIAKKNKASREVENELKAKLPNISFKYDQTYKVQGNKNTILVSPVIY 409
Db 462 --LVGMLRGIASGMKYLSDMGVYVHRDLAAR--NILIN-SNLVCKVSDFLGSLRVL----- 511
QY 410 TKANDNPDA 419
Db 511 -----EDDPEA 516

RESULT 9
US-08-316-397B-2
; Sequence 2, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-9880
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 859 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-316-397B-2
```

Query Match 5.0%; Score 107; DB 1; Length 859;
Best Local Similarity 22.0%; Pred. No. 0.23;

```
Matches 93; Conservative 55; Mismatches 137; Indels 138; Gaps 20;
QY 41 DSKPOAETLKH-KVSNDSIRIALTDPDPRWTSAAQDIISYVDETEAATSTITKNOAON 99
Db 152 DKEKAEFLKSAKQSPAGIIG-----NQIRTKQKFMGVFDE-----SLKERGEAK 198
QY 100 N-----WLTOQANLSPAPKGFIIAPENGSGVGTAVN-----TIADKGIP 138
Db 199 NGSGPTGDMW--DIFLS-----FIFDKOSSDVKEAINOEPLPHVQPDIASTTHIQGLP 251
QY 139 IVAYDRILIGSKYDWYVSDNEKVEGLOGL-----SLAAGLL 176
Db 252 PESRDLL-----DERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLHNHTLSSVLM 304
QY 177 GREDEGAFDSIDQWNEVYLKSHMPOETISFYTIA---GSODDNNQSYFY-----NGAMKVL 227
Db 305 GSHDG-----IEPKVSLLYAGNGGEGARHDMNATVGYKDOQGNVATII 349
QY 228 KELMKNQNKIIDLSPEGEN--AVYVPGWNYGTAGORISQFSLTINKDPAGGNKIKAVGSK 285
Db 350 NVHMKNSGGLVIAGGEGKINNPFSFYLYKEDQLTGSQRALSQEEIQ-----NKID----- 399
QY 286 PASIFKGLAPN-----DGMABQAITKLEGEFTQK-----IFVTRQDYND 327
Db 399 ----FMEFLAQNNAKLDSLSEKEKEKFRNEIKDFQKDSRPYLDALGNDRIAFYVSKKDKPKH 454
QY 328 KAK-TFIKGDQDNMTIYKPKDKVLGVAVVEVLRLVLIKKKASRESEVENELKAKLPNISEK 386
Db 455 SALITEFNKGDSYTL-----KVMGK---KQIKALDRKKNVTLOQNLKHGDMVFMVYSNEK 507
QY 387 YDN 389
Db 508 YTN 510

RESULT 10
US-08-316-397B-4
; Sequence 4, Application US/08316397B
; Patent No. 5733740
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TUMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SPRATT, GWENDOLYN D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 2200.030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-9880
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
```

;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-316-397B-4

Query Match 5.0%; Score 107; DB 1; Length 1181;
Best Local Similarity 22.0%; Pred. No. 0.38;
Matches 93; Conservative 55; Mismatches 137; Indels 138; Gaps 20;

QY 41 DSKPOAETLKH-KVSNDSIRIALTDPNPRWISAOKDIISYVDETEAATSTITKNQDAQN 99
DB 152 DDKERAEFLKSAKQSFAGIIG-----NQIRTDQKFMGVFE-----SLKREQEAEK 198
QY 100 N-----WLTQOANLSPAPKGFIIAPENGSGVGTAVN-----TIADKGIP 138
DB 199 NGGPTGGDWL--DIFLS-----FIFDKQSSDVKEAINOEPLPHVQPDITATSTTHIQGLP 251
QY 139 IVAYDRLTGSKDYWYVSFDNEKVGELQGL-----SLAAGLL 176
DB 252 PESRDLL-----DERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLHNHTLSSVLM 304
QY 177 GREDAFSDIQMNEVLKSHMPQETISFTIA---GSQDDNNSQYFY-----NGAMKYL 227
DB 305 GSHDG-----IEPEKVSLLYAGNGGFGARHDMNATVGYKDDQGNVATII 349
QY 228 KELMKNQNKIIIDLSPEGEN--AVYVPGWNYGTAGQRIQSFLTINKDPAGGNKIKAVGSK 285
DB 350 NVHMKNKSGLVIAAGGKGINNPSFYLYKEDQLTGSORALSQEEIQ-----NKID----- 399
QY 286 PASIFKGFILAPN-----DGMAEQAITKLEGGFTQK-----IFVTRQDYND 327
DB 399 -----FMEFLAQNNNAKLDLSLSEKEKEFRNEIKDFQKDSKPYPDLALGNDRIAFVSKKDPKH 454
QY 328 KAK-TFIKGDQDNMTIYKPKDVLGKVAVVEVLRLVLIKKNKASRSEVENELKAKLPNISFK 386
DB 455 SALITEFNKGDLSYTL-----KVMGK---KQIKALDREKNVTLOGLNKHGDMVFMVYSNFK 507
QY 387 YDN 389
DB 508 YTN 510

RESULT 11
US-09-034-306-2
; Sequence 2, Application US/09034306
; Patent No. 5876943
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TOMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,306
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,397

;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: SPRATT, GWENDOLYN D.
;; REGISTRATION NUMBER: 36,016
;; REFERENCE/DOCKET NUMBER: 2200.030
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 404/688-0770
;; TELEFAX: 404/688-9880
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 859 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-034-306-2

Query Match 5.0%; Score 107; DB 2; Length 859;
Best Local Similarity 22.0%; Pred. No. 0.23;
Matches 93; Conservative 55; Mismatches 137; Indels 138; Gaps 20;

QY 41 DSKPOAETLKH-KVSNDSIRIALTDPNPRWISAOKDIISYVDETEAATSTITKNQDAQN 99
DB 152 DDKERAEFLKSAKQSFAGIIG-----NQIRTDQKFMGVFE-----SLKREQEAEK 198
QY 100 N-----WLTQOANLSPAPKGFIIAPENGSGVGTAVN-----TIADKGIP 138
DB 199 NGGPTGGDWL--DIFLS-----FIFDKQSSDVKEAINOEPLPHVQPDITATSTTHIQGLP 251
QY 139 IVAYDRLTGSKDYWYVSFDNEKVGELQGL-----SLAAGLL 176
DB 252 PESRDLL-----DERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLHNHTLSSVLM 304
QY 177 GREDAFSDIQMNEVLKSHMPQETISFTIA---GSQDDNNSQYFY-----NGAMKYL 227
DB 305 GSHDG-----IEPEKVSLLYAGNGGFGARHDMNATVGYKDDQGNVATII 349
QY 228 KELMKNQNKIIIDLSPEGEN--AVYVPGWNYGTAGQRIQSFLTINKDPAGGNKIKAVGSK 285
DB 350 NVHMKNKSGLVIAAGGKGINNPSFYLYKEDQLTGSORALSQEEIQ-----NKID----- 399
QY 286 PASIFKGFILAPN-----DGMAEQAITKLEGGFTQK-----IFVTRQDYND 327
DB 399 -----FMEFLAQNNNAKLDLSLSEKEKEFRNEIKDFQKDSKPYPDLALGNDRIAFVSKKDPKH 454
QY 328 KAK-TFIKGDQDNMTIYKPKDVLGKVAVVEVLRLVLIKKNKASRSEVENELKAKLPNISFK 386
DB 455 SALITEFNKGDLSYTL-----KVMGK---KQIKALDREKNVTLOGLNKHGDMVFMVYSNFK 507
QY 387 YDN 389
DB 508 YTN 510

RESULT 12
US-09-034-306-4
; Sequence 4, Application US/09034306
; Patent No. 5876943
; GENERAL INFORMATION:
; APPLICANT: COVER, TIMOTHY L.
; APPLICANT: BLASER, MARTIN J.
; APPLICANT: HARRY KLEANTHOUS
; APPLICANT: TOMMURU, MURALI K.R.
; TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
; TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/09/034,306
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-034-306-4

Query Match 5.0%; Score 107; DB 2; Length 1181;
Best Local Similarity 22.0%; Pred. No. 0.38;
Matches 93; Conservative 55; Mismatches 137; Indels 138; Gaps 20;

QY 41 DSKPOAETLKH-KVSNDSIRIALTPDNPRTISAQKDIISYDETEAATSTITKNQDAQN 99
DB 152 DKEKAEFLKSAQSFAGIIG-----NQRTQKFMGVFDE-----SLKREQAEK 198
QY 100 N-----WLTOQANLSPAPKGFIIAPENGSGVGTAVN-----TIADKGIP 138
DB 199 NGGPTGDMWL--DIFLS-----FIFDKQSSDVKEAINQEPHVPQDIATSTTHIOGLP 251
QY 139 IVAYDRLLTSGDKYDNYVFDNEKVGELQGL-----SLAAGLL 176
DB 252 PESRDL-----DERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLIHNTLSVLM 304
QY 177 GKEDGAFDSIDQMEYLNKSHMPOETISFYTIA---GSODDNNISQFY-----NGAMKYL 227
DB 305 GSHDG-----IEPEKVSLLYAGNGGFGAKHDWATVGYKDOOGNNVATII 349
QY 228 KEMKNSQNKIIDLSPEGEN--AVYVPGWNYGTAGRIQSFLTINKDPAGNKKIKAVGSK 285
DB 350 NVHMKNKSGLVIAAGEKGINNPSFYLYKEDQLTGSORALSQEEIQ-----NKID----- 399
QY 286 PASIEKGFAPN---DGMAEQAITKLKLGFTQK-----IFVTRODYND 327
DB 399 -----FMEFLAQNNAKLDSUSEKEKEFRNEIKDFQKDSKPYLDALGNDRIFAVSKDKPH 454
QY 328 KAK-TFIKDGQDNMTYKPKVLGKVAVEVLRLVLIKAKNSRSEVENELKAKLPNISFK 386
DB 455 SALITEFNKGDLSYTL---KVWKG---KQIKALDREKNVTLOGNLKHDGVFMVFNYSNEK 507
QY 387 YDN 389
DB 508 YTN 510

RESULT 13
PCT-US93-09782-2
Sequence 2, Application PC/TUS9309782
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TUMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING

TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-09782-2

Query Match 5.0%; Score 107; DB 3; Length 859;
Best Local Similarity 22.0%; Pred. No. 0.23;
Matches 93; Conservative 55; Mismatches 137; Indels 138; Gaps 20;

QY 41 DSKPOAETLKH-KVSNDSIRIALTPDNPRTISAQKDIISYDETEAATSTITKNQDAQN 99
DB 152 DKEKAEFLKSAQSFAGIIG-----NQRTQKFMGVFDE-----SLKREQAEK 198
QY 100 N-----WLTOQANLSPAPKGFIIAPENGSGVGTAVN-----TIADKGIP 138
DB 199 NGGPTGDMWL--DIFLS-----FIFDKQSSDVKEAINQEPHVPQDIATSTTHIOGLP 251
QY 139 IVAYDRLLTSGDKYDNYVFDNEKVGELQGL-----SLAAGLL 176
DB 252 PESRDL-----DERGNFSKFTLGDMEMLDVEGVADMDPNYKFNQLLIHNTLSVLM 304
QY 177 GKEDGAFDSIDQMEYLNKSHMPOETISFYTIA---GSODDNNISQFY-----NGAMKYL 227
DB 305 GSHDG-----IEPEKVSLLYAGNGGFGAKHDWATVGYKDOOGNNVATII 349
QY 228 KEMKNSQNKIIDLSPEGEN--AVYVPGWNYGTAGRIQSFLTINKDPAGNKKIKAVGSK 285
DB 350 NVHMKNKSGLVIAAGEKGINNPSFYLYKEDQLTGSORALSQEEIQ-----NKID----- 399
QY 286 PASIEKGFAPN---DGMAEQAITKLKLGFTQK-----IFVTRODYND 327
DB 399 -----FMEFLAQNNAKLDSUSEKEKEFRNEIKDFQKDSKPYLDALGNDRIFAVSKDKPH 454
QY 328 KAK-TFIKDGQDNMTYKPKVLGKVAVEVLRLVLIKAKNSRSEVENELKAKLPNISFK 386
DB 455 SALITEFNKGDLSYTL---KVWKG---KQIKALDREKNVTLOGNLKHDGVFMVFNYSNEK 507
QY 387 YDN 389
DB 508 YTN 510

RESULT 14
PCT-US93-09782-4

Sequence 4, Application PC/TUS9309782
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TUMURU, MURALI K.R.
TITLE OF INVENTION: THE TGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-09782-4

Query Match 5.0%; Score 107; DB 3; Length 1181;
Best Local Similarity 22.0%; Pred. No. 0.38;
Matches 93; Conservative 55; Mismatches 137; Indels 138; Gaps 20;
QY 41 DSKPQATLKH-KVSNDSIRIALTPDNPWISAQKDIISYVDETEAATSTITKNQDAQN 99
DB 152 DKEAEFLKSAQSFAGIIG-----NQRTDQKFGVFE-----SLKEREAEK 198
QY 100 N-----WLTQANLSPAPKGFIIAPENGSGVGTAVN-----TIADKGP 138
DB 199 NGGPTGDWL--DIFLS-----FIFDKQSSDVKEAINQEPPLPHVQPDIASTTHIQGLP 251
QY 139 IVAYDRLLTGSDKYVYFSDNEKVGELGL-----SLAAGLL 176
DB 252 PESRDL-----DERGNFSKFTLGDMEMLDVEGVADMDPNKFNQLLIHNNLSVLM 304
QY 177 GKEDGAFDSIDOMNEYLKSHMPOETISFYTIA-----GSODDNNNSOYFY-----NGAMKYL 227
DB 305 GSHDG-----IPEKYSLLYAGNGGFGAKHDNATVGYKDOQGNVATII 349
QY 228 KMLNKSQNKIIDLSPEGEN--AVYVPGWNYGTAGRIOSFLTINKDPAGGNKIRAVGSK 285
DB 350 NVHKNNGSLVIAAGGKGINNPSFYLYKEDQLTGSQALSQEBIQ-----NKID----- 399
QY 286 PASIFKGFAPN-----DGAEQAITKLEGFDTQK-----IFVTROYND 327
DB 399 -----FMEFLAQNNNAKLDLSSEKEKEFRNEIKDFQDSKPYLDALGNDRIFAFTSKKDPKH 454
QY 328 KAK-TRIKDGDQNTLYKPKDVLGKVAVEVLRVLIKKNKASSEVENELKAKLPNISEK 386
DB 455 SALITEFNKGDLSYTL-----KVMGK-----KQIKALDREKNVTLQGNLKHGDMVFNYSNEK 507
QY 387 YDN 389

DB 508 YTN 510
RESULT 15
US-08-409-995-4
Sequence 4, Application US/08409995
Patent No. 5646259
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen I.
APPLICANT: St. Geme III, Joseph W.
TITLE OF INVENTION: Haemophilus Adhesion Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

Query Match 5.0%; Score 106.5; DB 1; Length 1912;
Best Local Similarity 19.6%; Pred. No. 0.91;
Matches 91; Conservative 64; Mismatches 197; Indels 113; Gaps 19;
QY 35 ESGSTSDSK--PQAE TLKH-----KVSNDISIRI-----ALTDPNRPWISAQKDIISY 80
DB 259 ENAKTEVKFTPKTSVIKEKDGKLTGKENDNTKVTSTATDNTDEGNGLVTAKAVIDA 318
QY 81 VDTEAATSTITKNQDAQNWLTPQANLSPAPKGFIIAPENGSGVGTAVNTIADKGIPIV 140
DB 319 VNKAGRVKTTTAN-----GQNGDFATVAGTNTVTFESDGTTSVTKDTNGNGIIV 370
QY 141 AYDRLLTGSDKYVYFSDNEKVGELGSLAAGL--LGKE----- 180
DB 371 KYDAKVGDKLFD-----SDKTIADTTALTVTGGKVAETAKEDDKKKLVNAGDLVTALGN 426
QY 180 -----DGAFDSIDOMNEYLKSHMPOETISF-----YTIAGSODDNNNSOYFYNGAMK 225
DB 427 LSWKAKAEADTDGALEGISKDQEVKAG-----ETVTFKAGKNLKVKODGANFTYISLODAUT 482
QY 226 VLKELM-----KNSQNKIIDLSPEGENAVYVPGWNYGTAGRIQSFLTINKD--PAGE 276
DB 483 GLTSITLGTGTNGNDKATVIN---KDLGTTTPAGNGGTG---TNTISVTKDGKICAGN 535
QY 277 NIKKAVGSKPAPISFGFLAPNDGMAEQAITKULLEGFDQKIFVTROYNDKAKTFIKD- 336
DB 536 KAITNV-----ASGLRAYDDANFDVNLNNSATDLNRHVEDAYKGLLNILNEKNANKQPLVTD 591

Qy 336 -----GD-----QNTIYKPKVLGKVAEVL-----RVLIAKKNKSRSEVENE 375
Db 592 TAATVGDRLKLGWVSTKNGTKESNOV--KQADEVLFTGAGAATVTSKSENGKHTTIVS 649
Qy 376 LKAKLPNISFKYDNOTYKVOGKNINTILVSPV-----IVTKANVD 415
Db 650 VAETKADCGLEKGDGTIKLKVDNONTDNLTVGNNGTAVTKGGFE 694

Search completed: November 13, 1999, 10:56:26
Job time: 1355 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 02:07:34 ; Search time 251.81 Seconds
(without alignments)
105.336 Million cell updates/sec

Title: US-08-913-430-2

Perfect score: 2143

Sequence: 1 MKKMLRKFLYSAYATSL.....NTILSPVIVTKANYNDPDA 419

Scoring table: BLOSUM62

Searched: 497124 seqs, 53304441 residues

Database : Pending_Patents_AA:*
1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2143	100.0	419	13	US-08-913-430-2
2	2129	99.3	419	4	US-08-183-774-1
3	275.5	12.9	363	15	US-09-107-532-5091
4	138	6.4	29	13	US-08-913-430-4
5	122.5	5.7	433	15	US-09-107-532-5822
6	122	5.7	277	14	US-09-071-035-30
7	122	5.7	278	15	US-09-134-000-4965
8	119	5.6	25	13	US-08-913-430-3
9	118	5.5	643	10	US-08-600-993A-47
10	117.5	5.5	390	15	US-09-107-532-6425
11	116.5	5.4	654	14	US-09-090-793-2
12	116.5	5.4	654	16	US-09-231-899-2
13	115.5	5.3	307	1	PCT-US97-22578-203
14	114.5	5.3	3351	12	US-08-827-356-5564
15	114	5.3	930	16	US-09-200-650-3
16	113.5	5.3	2353	13	US-08-913-942-4
17	113.5	5.3	2411	16	US-09-268-347-36
18	113.5	5.3	2354	16	US-09-268-347-47
19	113.5	5.3	306	16	US-09-296-284-6

20	113	5.3	658	16	US-09-248-796-17674	Sequence 17674, A
21	113	5.3	658	19	US-60-096-409-17674	Sequence 17674, A
22	112.5	5.2	3135	1	PCT-US94-00547-2	Sequence 2, Appli
23	112.5	5.2	3135	6	US-08-323-170-2	Sequence 2, Appli
24	112.5	5.2	3135	6	US-08-323-170A-2	Sequence 2, Appli
25	112.5	5.2	3135	13	US-08-954-441-2	Sequence 2, Appli
26	112.5	5.2	1315	16	US-09-200-650-5	Sequence 5, Appli
27	110.5	5.2	931	16	US-09-252-088-26	Sequence 26, Appli
28	109.5	5.1	994	5	US-08-235-407-2	Sequence 2, Appli
29	109.5	5.1	994	9	US-08-542-635-2	Sequence 2, Appli
30	109.5	5.1	994	11	US-08-727-463-2	Sequence 2, Appli
31	109.5	5.1	994	11	US-08-727-463-2	Sequence 2, Appli
32	109.5	5.1	994	11	US-08-730-700-2	Sequence 2, Appli
33	109.5	5.1	994	11	US-08-730-700A-2	Sequence 2, Appli
34	109.5	5.1	1221	15	US-09-107-532-3959	Sequence 3959, Ap
35	109.5	5.1	2314	16	US-09-268-347-49	Sequence 49, Appl
36	109.5	5.1	280	16	US-09-296-284-27	Sequence 27, Appl
37	109	5.1	392	16	US-09-248-796-14809	Sequence 14809, A
38	109	5.1	392	19	US-60-096-409-14809	Sequence 14809, A
39	108.5	5.1	577	17	US-09-328-352-7698	Sequence 7698, Ap
40	108	5.0	2142	19	US-60-128-476-4493	Sequence 4493, Ap
41	108	5.0	319	23	US-09-417-507-37952	Sequence 37952, A
42	107	5.0	3829	12	US-08-827-356-5567	Sequence 5567, Ap
43	106.5	5.0	1912	10	US-08-685-467-4	Sequence 4, Appli
44	106	4.9	1098	10	US-08-685-467-2	Sequence 2, Appli
45	106	4.9	796	12	US-08-827-356-3416	Sequence 3416, Ap

ALIGNMENTS

RESULT 1
US-08-913-430-2
; Sequence 2, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; EARLIER FILING DATE: 1997-12-09
; EARLIER FILING DATE: 1996-03-15
; EARLIER FILING DATE: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0 - beta
; SEQ ID NO 2
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-2

Query Match	100.0%	Score	2143	DB	13	Length	419
Best Local Similarity	100.0%	Pred. No.	4.8e-171				
Matches	419	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
QY	1	MKKMLRKFLYSAYATSLASIIAFVAAAGCGTSGTSDSKPQATLKHVNSDSIRI	60				
Db	1	MKKMLRKFLYSAYATSLASIIAFVAAAGCGTSGTSDSKPQATLKHVNSDSIRI	60				
QY	61	ALTDPPNPWISAKDIIISYVDETEATSTIYNQDQANNLTQANLSPAPKGFIIAE	120				
Db	61	ALTDPPNPWISAKDIIISYVDETEATSTIYNQDQANNLTQANLSPAPKGFIIAE	120				
QY	121	NSGSGVTANTADKGIPIVAYDRLLITGSKYDWTYVFNKVGELQGLSLAAGLLGKD	180				
Db	121	NSGSGVTANTADKGIPIVAYDRLLITGSKYDWTYVFNKVGELQGLSLAAGLLGKD	180				
QY	181	GAFDSIDQNEYLKSHMPQETISFYTIAGSQDDNNSQYFNGAMKYLKELMKNQNKIID	240				
Db	181	GAFDSIDQNEYLKSHMPQETISFYTIAGSQDDNNSQYFNGAMKYLKELMKNQNKIID	240				

Db 181 GAFSDIDOMNEYLKSHMPOETISFYTIAGSODDNNNSQYFYNGAMKVLKELMKNSQNKIID 240
QY 241 LSPGEGNAVVPVGNNGYTAGORIOSFLTINKDPAGGNKIRAVGSKPASIFKGFAPNDGM 300
Db 241 LSPGEGNAVVPVGNNGYTAGORIOSFLTINKDPAGGNKIRAVGSKPASIFKGFAPNDGM 300
QY 301 AEQAITKLKLEGFTQKIFVTRQDYNDAKTFIKDGDQNMNTIYKPKDVLGKVAVEVLRLV 360
Db 301 AEQAITKLKLEGFTQKIFVTRQDYNDAKTFIKDGDQNMNTIYKPKDVLGKVAVEVLRLV 360
QY 361 IAKKNKASRSEVENELKAKLPNISFKYDQNTYKVOGKNINTILVSPVIVTKANVDNPD 419
Db 361 IAKKNKASRSEVENELKAKLPNISFKYDQNTYKVOGKNINTILVSPVIVTKANVDNPD 419

RESULT 2
US-08-183-774-1
; Sequence 1, Application US/08183774
; GENERAL INFORMATION:
; APPLICANT: SETO, Yasuhiro
; APPLICANT: FUTO, Satoshi
; APPLICANT: MITSUSE, Shizuo
; APPLICANT: MATSUO, Kanako
; APPLICANT: TSUNA, Mika
; TITLE OF INVENTION: DNA's Encoding Surface Antigen of
; TITLE OF INVENTION: Mycoplasma hyopneumoniae, DNA Fragments for Primer,
; TITLE OF INVENTION: Recombinant Antigenic Peptides and Diagnostic Method of
; TITLE OF INVENTION: Mycoplasma Pneumoniae of Swine Using Same
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/183,774
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 07/747,015
; FILING DATE: 19-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, Norman F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4183-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEFAX: (703)486-2347
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 419 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-183-774-1

Query Match 99.3%; Score 2129; DB 4; Length 419;
Best Local Similarity 99.5%; Pred. No. 7.2e-170;
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKMLRKFLYSSAIYATSLAIIAFVAAAGCGQTESGSDSKPQAEFLKHKVNSDSIRI 60
Db 1 MKMLRKFLYSSAIYATSLAIIAFVAAAGCGQTESGSDSKPQAEFLKHKVNSDSIRI 60

QY 61 ALTDPDNPRWISAQKDIISVDETEAATSTITTKNOAQNNWLTQQANLSPAPKGFIIAPE 120
Db 61 ALTDPDNPRWISAQKDIISVDETEAATSTITTKNOAQNNWLTQQANLSPAPKGFIIAPE 120
QY 121 NGSGVGTAVNTIADKGIPIVAYDRLITGSDKYDWYVSFDNEKYGELQGLSLAAGLLGLE 180
Db 121 NGSGVGTAVNTIADKGIPIVAYDRLITGSDKYDWYVSFDNEKYGELQGLSLAAGLLGLE 180
QY 181 GAFSDIDOMNEYLKSHMPOETISFYTIAGSODDNNNSQYFYNGAMKVLKELMKNSQNKIID 240
Db 181 GAFSDIDOMNEYLKSHMPOETISFYTIAGSODDNNNSQYFYNGAMKVLKELMKNSQNKIID 240
QY 241 LSPGEGNAVVPVGNNGYTAGORIOSFLTINKDPAGGNKIRAVGSKPASIFKGFAPNDGM 300
Db 241 LSPGEGNAVVPVGNNGYTAGORIOSFLTINKDPAGGNKIRAVGSKPASIFKGFAPNDGM 300
QY 301 AEQAITKLKLEGFTQKIFVTRQDYNDAKTFIKDGDQNMNTIYKPKDVLGKVAVEVLRLV 360
Db 301 AEQAITKLKLEGFTQKIFVTRQDYNDAKTFIKDGDQNMNTIYKPKDVLGKVAVEVLRLV 360
QY 361 IAKKNKASRSEVENELKAKLPNISFKYDQNTYKVOGKNINTILVSPVIVTKANVDNPD 419
Db 361 IAKKNKASRSEVENELKAKLPNISFKYDQNTYKVOGKNINTILVSPVIVTKANVDNPD 419

RESULT 3
US-09-107-532-5091
; Sequence 5091, Application US/09107532
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7308
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085598
; FILING DATE: May 14, 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5091:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE: NAME/KEY: misc_feature

LOCATION: 1...363
US-09-107-532-5091

```

Query Match          12.9% Score 275.5; DB 15; Length 363;
Best Local Similarity 25.4%; Pred. No. 9e-15;
Matches 108; Conservative 65; Mismatches 158; Indels 95; Gaps 17;

QY 1 MKKMLRKFLYSSAIYATSLASIIATFAAGCGQTGESGTSQKPAETLKHVKVNSDSTRI 60
   || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 MKNKFRK-----IVGTIIAILTASILLACG--NSGSADD-----STGYVGI 40
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 ALTPDPNPRISAOAKDIISVYDETEAAATSTTKNQDAQNNWLTOQANLSPAPKG---FII 117
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 41 AMPTKSAERWIADGNMNMVSELEKLGKGYTD-LQYGEDKVENOVAQIENM--ITKGVDTLVI 97
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 118 APENGSGVGTAVNTIADGPIVADYDRLITGSDKYDMVYSPDNKVGSELOGLSLAAGLLG 177
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98 ASIDGSALTDVLAKAKEADIKVIAYDRLLMNSENVYIATEDNFGV---GVSQAA----- 150
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 178 KEGAFDSIDOMEYLSKHPM-QETISFYIT---AGSODDNNSOYFYNGAMKVLKELMKN 233
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 150 -----YIEHLGLKEGKGPFTELEFGGSPDDNNALINYNMGVMSVLQPYMDN 195
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 234 SONKIIDLSPGENA---VVYPGWNYSAGORIOSFETINKDPAGGNKIKKAVGSKPASIF 290
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 196 GOLVV---PSGQTSFSQIATLRWDGSTAQARMNLLISANT---DOTLDV----- 241
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 291 KGFLAPDGMGAEOAITKLLEGFTQK---IFVTRQYNDKAKTFIKDGDQNMWITYPKDK 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 ---LSPYDPISLGIISLKGVGYSSEKPLVITGQDATVAGVKSITAGEQTQIFKDR 297
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 348 VLGKVAVEVLRLVLIANKKASRSEVENELKAKLPNISFKYDNOTYKVQGNINTILYSPV 407
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 298 ILAKNTIEMIKAI-----SDGEEVPVN-----DTETYDNGVKTVPYLANTV 339
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 408 IVTKAN 413
   | | | |
Db 340 SVDKDN 345

RESULT 4
US-08-913-430-4
; Sequence 4, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 4
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-4

```

```
Query Match      6.4%; Score 138; DB 13; Length 29;
Best Local Similarity 100.0%; Pred. NO. 5.4e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RESULT 5
US-09-107-532-5822
; Sequence 5822, Application US/09107532
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7308
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085598
; FILING DATE: May 14, 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5822:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...433
US-09-107-532-5822

```

Query Match      5.7%; Score 122.5; DB 15; Length 433;
Best Local Similarity 19.6%; Pred. No. 0.077;
Matches 96; Conservative 73; Mismatches 178; Indels 143; Gaps 23;

QY      2  KKMLRKKFLYSSAIYATSLASIIAFVAAAGCGQTESGTD-----SKPOAETLK--- 51
      |||  |||  :  :  :  |||  :  |||  :  :  :  :  :  :  :  :  :  :  :  :
Db      5  KKM---KKLIIVTCLLAI-----GFLITAGESOSSNESQOITTIQFVHSSVEQRLSVIN 56

QY      51  -----HKVNSNDRIALTDPDN-----PRWISAQKDIIISYVDTEAATS 89
      |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      57  ELVADFEKENPDIIKQVPVEESAFNTKIYTLARSGKLPEVMEYSQDFAKYMDKDELIDQ 116

QY      90  TITKN---QDAONNLWTLQANL---SPAPKGFILAPENG---SGVGTAVNTADKGPIVAY 142
      |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      117  TAVONVTEEAGEDNYDGAKNLVRSEDGKSYIAAPIGTVQGIWYDKEKLEAGF----- 172

QY      143  DRLITGSDKYDW-----YVSFDNKEKVGELQ-----GLSLAAGLLGK 178
      |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      172  -----SEPNWDDILKIAHQFTDCKENKQYGIAMPTAEGTMSQAFSOFALNSGANVLDD 225

```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 11:33:28 ; Search time 75.45 Seconds
(without alignments)
222.498 Million cell updates/sec

Title: US-08-913-430-2
Perfect score: 2143
Sequence: 1 MKMLRKKFLYSAYATSL.....NTILSPVITVKANVDNPD 419

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2136	99.7	419	2	A56153	46K surface antige
2	326	15.2	332	2	C64183	xyf protein - Hae
3	305.5	14.3	330	2	S47787	xyf protein - Esc
4	171	8.0	170	2	A43654	probable periplasm
5	133.5	6.2	305	2	A69690	ribose ABC transpo
6	124	5.8	630	2	JQ1670	polygalacturonase
7	123	5.7	332	2	S15554	D-galactose-bindin
8	121	5.6	271	1	JGEBRT	D-ribose-binding p
9	118	5.5	318	2	S56453	hypothetical 32.1K
10	118	5.5	643	2	S54395	permease-like prot
11	117.5	5.5	296	1	JGECR	periplasmic ribose
12	116.5	5.4	292	2	A64073	D-ribose-binding p
13	116	5.4	332	1	JGECG	D-galactose-bindin
14	113.5	5.3	1119	2	B70126	surface-located me
15	113	5.3	332	2	S29390	galactose-binding
16	112.5	5.2	3135	2	A48584	transmission block
17	111.5	5.2	1451	2	A36468	spt6 protein - yea
18	111	5.2	1104	2	A36866	microbial collagen
19	111	5.2	3016	2	S77300	hypothetical prote
20	110.5	5.2	345	2	S72333	periplasmic bindin
21	108.5	5.1	586	2	JC4778	pectinesterase (EC
22	108.5	5.1	451	2	E70013	hypothetical prote
23	108	5.0	1302	2	JC6009	surface-located me
24	107	5.0	349	2	G64096	D-galactose-bindin
25	106.5	5.0	802	1	YU2NS	sucrose synthase (
26	106.5	5.0	315	2	S32330	ferrichrome ABC tr
27	106.5	5.0	688	2	D70129	hypothetical prote
28	106.5	5.0	1225	2	B64234	hypothetical prote
29	106	4.9	683	2	S69780	outer membrane pro
30	105.5	4.9	2491	2	A57036	talin - slime mold
31	104.5	4.9	334	2	T03556	probable D-ribose
32	104.5	4.9	718	2	E71100	hypothetical prote
33	104	4.9	1518	2	S37328	probable purine nu
34	103.5	4.8	1407	1	BWBXM1	TATA box-binding p
35	103.5	4.8	460	2	S35772	translation elonga
36	103	4.8	876	2	D70971	hypothetical prote
37	103	4.8	591	2	D64204	membrane lipoprote
38	103	4.8	697	2	S73428	probable lipoprote
39	102.5	4.8	991	2	I78843	receptor protein-t

40 102 4.8 4199 2 S76412 hypothetical prote
41 102 4.8 641 2 S69651 SAC2 protein - yea
42 101.5 4.7 1165 2 A70423 valyl-tRNA synthet
43 101 4.7 762 2 A34355 cell surface prote
44 101 4.7 348 2 B69790 hypothetical prote
45 101 4.7 1155 2 G64332 FUN12/bif-2 family

ALIGNMENTS

RESULT 1
A56153
46K surface antigen precursor - Mycoplasma hyopneumoniae (SGC3)
C:Species: Mycoplasma hyopneumoniae
C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 09-Sep-1997
C:Accession: A56153
R:Puto, S.; Seto, Y.; Mitsuse, S.; Mori, Y.; Suzuki, T.; Kawai, K.
J. Bacteriol. 177, 1915-1917, 1995
A:Title: Molecular cloning of a 46-kilodalton surface antigen (P46) gene from Mycopla
A:Reference number: A56153; MUID:95204368
A:Accession: A56153
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <PUT>
A:Cross-references: GB:D16682; NID:g517134; PID:d1004604; PID:g1220114
C:Genetics:
A:Gene: P46
A:Genetic code: SGC3
C:Keywords: surface antigen

Query Match 99.7%; Score 2136; DB 2; Length 419;
Best Local Similarity 99.8%; Pred. No. 3e-136;
Matches 418; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKMLRKKFLYSAYATSLAIAFAAGCGQTSGSTSDSKPQAEATLKHKVSNDISRI 60
Db 1 MKMLRKKFLYSAYATSLAIAFAAGCGQTSGSTSDSKPQAEATLKHKVSNDISRI 60
QY 61 ALTDPPNPWISAKDIIISVDETEAATITKNDQAQNWLTTQANLSPAPKGFIIAPE 120
Db 61 ALTDPPNPWISAKDIIISVDETEAATITKNDQAQNWLTTQANLSPAPKGFIIAPE 120
QY 121 NSGSGVTAVNTADKGIPIVAVDRITGSKDYVDFNEKVGELQGLSLAAGLLGKED 180
Db 121 NSGSGVTAVNTADKGIPIVAVDRITGSKDYVDFNEKVGELQGLSLAAGLLGKED 180
QY 181 GAFDSIDQMNVEYLKSHMPQETISFYIAGSQDDNNSQYFYNGAMKVKELMKNSQNKIID 240
Db 181 GAFDSIDQMNVEYLKSHMPQETISFYIAGSQDDNNSQYFYNGAMKVKELMKNSQNKIID 240
QY 241 LSPGEGNAVYVPGWNYGTAGRIQSFLLTINKDPAGNKKIKAVGSKPASIFKGLAPNDGM 300
Db 241 LSPGEGNAVYVPGWNYGTAGRIQSFLLTINKDPAGNKKIKAVGSKPASIFKGLAPNDGM 300
QY 301 AEQAITKLKLEGFDTKIFVTROYNDKAKTFTKDCDQNNITKPKDKVLGKVAVEVLVL 360
Db 301 AEQAITKLKLEGFDTKIFVTROYNDKAKTFTKDCDQNNITKPKDKVLGKVAVEVLVL 360
QY 361 IAKKNKASRSEVENELKALPNISFYDKYQKQKNINTILVSPVITVKANVDNPD 419
Db 361 IAKKNKASRSEVENELKALPNISFYDKYQKQKNINTILVSPVITVKANVDNPD 419

RESULT 2
C64183
xyf protein - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Oct-1997
C:Accession: C64183
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman

A; Experimental source: strain K-12, substrain MG1655
C; Genetics:

```

Db      256 ASIKDRKLSATVAQQPELIGKLTEAADDILHGK-----KVQKTISAPL 299

```

RESULT 6

C;Species: *Lycopersicon esculentum* (tomato)
C;Date: 10-Jun-1993 #sequence revision 10-Jun-1993 #text change

Plant Cell 4, 1147-1156, 1992

A;Accession: JQ1670
A:Message: tyne. mRNA

A;Cross-references: GB:M98466; NID:GI/04/9; PID:GI/0480
A;Accession: PQ0447

A; Experimental source: fruit, cv. Ailsa Craig

F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-108/Domain: amino-terminal propeptide #status predicted <PR>

F; 124, 142, 256, 334, 369, 387/Binding site: carbohydrate (Asn) (cova

Query Match 5.8%; Score 124; DB 2; Length 630;
Best Local Similarity 22.7%; pred No 0 63.

QY 20 LASIIAFVAGCGQTEGSTSDSKPQ-AETLKHVNSDSIRIALTDPDNPRWIS

Q: 70 CYNTHIA ANN MC - TEMPORE - - ON THE EVENING OF NOVEMBER 1970

Db 65 SPLNAAQYATYTKLVADQNALTTLQHTFCSSANLMCAPD---LSPSLEKHSGDI

[illegible]

QY 167 QGLSLA-AGLLGKEDGAFDSIDQMNEYLKSHMPQET--ISFYTIAGSQDDNNSQ

QY 224 MKVLKELMKNSONKIIDLSPEGENAVYVPGWNYGTAGORIOSELTINKDPAGNN

238
 239
 240

QY 283 ---GSKPASIFKGF LAPNDGM AEQAIT K L E G F D T Q K I F V T R Q D Y N D K A K - -
 ! : | : ! : | : ! : |

D**b** 282 YGNGCNVPENHFTNYGAGNGPSE-----TFNSYRDQSNVGV

[illegible]

QY 387 -----YDNQT^{YKV}-----QGKNINTILVSP 406
| | . | . |
| | . | . |

Db 367 GVNNTFKDYVKDTATFSNYHNKTSQVLASLMEVNGGKKVNNRWVEP 412

RESULT 7
C1555A

C;Species: Citrobacter freundii
C.Date: 20-Feb-1995 #sequence revision 20-Feb-1995 #text change

```
C:\date: 20-Feb-1993 #sequence_revision 20-Feb-1993 #text_change
```


C;Accession: S15554
R;Galindo, R.L.; Dagget Garvin, L.; Hardies, S.C.
submitted to the EMBL Data Library, May 1991
A;Reference number: S15554
A;Accession: S15554
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-332 <GAL>
A;Cross-references: EMBL:X59389; NID:g40472; PID:g40473
C;Genetics:
A;Gene: mg18
C;Superfamily: D-galactose-binding protein
C;Keywords: calcium binding; chemotaxis; periplasmic space; sugar transport
F:37,114,175,177,181,234,259,279/Binding site: galactose (Asp, Asn, His, Asp, Arg, Asn,
Query Match 5.7%; Score 123; DB 2; Length 332;
Best Local Similarity 21.2%; Pred. No. 0.3;
Matches 85; Conservative 69; Mismatches 155; Indels 92; Gaps 20;
QY 5 LRKFLYSALYATSLASIIAFVAGCGQTESGSDSKPQAEITLKHVSNDSRIALTD 64
DB 1 MNKVLILSAV-----MASLFGAAHAADTRIGVTI-----YKIDNFMVVRK- 46
QY 65 PDNPRWISAOKDIISYVDETEAATSTITKNQDQANNWLTOQAN----LSPAPKGFIIAPE 120
DB 46 -----AIKED-----AKAPDVQLLMNDSQDQSKNDQIDVLLAKGVKALAINLV 91
QY 121 NSGSGVTAINTADKGIPIVAYDRILIT--GSDKYD--WYVSFDNEKVGELQGLSLAAGLL 176
DB 92 DPAAAGTVIEKARGQNPVIFVFNKPSRKALDSYKAYVGTDSKESGIQG-----DLI 146
QY 177 GKEDGAPDSIDQMEYLKSHMPQETISPTIAG--SQDDNNSQYFYNGAMKVLKEL---- 231
DB 147 AKHWAANPNWD-LNKDGK-----IQEVLKGEFGHPDAERTY-----VIELNDKG 193
QY 231 MNKSNQKIIDLSPGENAVYVPGNYSAGQRIQSFITKNDPAGGNKIKAVGSKPASIF 290
DB 194 IKTEQLQL-----DTAM-----WDTAQAKDKMDAAMS-----GPNANKIEV----- 231
QY 291 KGFAPNDGMAEQATKLEGFDTQKIFVTRQDYNDKAKFTFKDQDNMTIYKPKVLG 350
DB 231 ---TANNNDAMAGAVEALKAHNKTSVPVFGV--DALPEALALVKSGAMAGVFLNDANQA 285
QY 351 KVAVEVLRLVLIKAKNSRS--EVENELKAKLPNISFKYDN 389
DB 286 KATFDLAKNLADGGAADGTWKNKI-KVVPYGVGDKDN 325
RESULT 8
JGEBRT
D-ribose-binding protein - Salmonella typhimurium (tentative sequence)
C;Species: Salmonella typhimurium
C;Date: 25-Feb-1985 #sequence_revision 10-Feb-1995 #text_change 22-Nov-1996
C;Accession: A03426
R;Buckenmeyer, G.K.; Hermodson, M.A.
J. Biol. Chem. 258, 12957, 1983
A;Title: The amino acid sequence of D-ribose-binding protein from Salmonella typhimurium
A;Reference number: A03426; MUID:84032514
A;Accession: A03426
A;Molecule type: protein
A;Residues: 1-271 <BUC>
A;Experimental source: strain ST1
A;Note: only cyanogen bromide peptides differing in composition from the Escherichia coli
sequence shown
C;Superfamily: This periplasmic binding protein is involved in the high-affinity D-ribose me
C;Keywords: chemotaxis; lac repressor
F:1-271/Product: D-ribose-binding protein #status experimental <MAT>
F:13,89,90,141,190,215,235/Binding site: beta-D-ribofuranose (Asn, Asp, Arg, Asn, A
Query Match 5.6%; Score 121; DB 1; Length 271;

Best Local Similarity 22.7%; Pred. No. 0.3;
Matches 73; Conservative 47; Mismatches 128; Indels 74; Gaps 13;
QY 56 DSIRIALTDPDNPRWISAOKDIISYVDETEAATSTITKN--ODAQNWLITQOANLSP-- 111
DB 2 DTIALIVSTLNNPFFVSLK-----DGAQKEADKLGYNLVLDSDQNNPAKELANVODLT 54
QY 111 --APKGFIIAPENGSGVGTAVNTIADKGIPIVAYDRILITGSDKYDWTYVSPDNKVGSELOQ 168
DB 55 VRGKILLINPTSDAVGNVAKNAQAKIPVITLDRAQTKGDVVS-HIASDNVLGGKIAG 113
QY 169 --LSLAAGLKGEDGAPDSIDQMEYLKSHMPQETISFYTIAGSDQNNNSQYFYNGAMKV 226
DB 114 DYIAKRAKAG-----EGA-----KVTELQGIAGTSAAREREGEQOAVAA 151
QY 227 LKELMNSQNKIIDLSPGENAVYVPGNYSAGQRIQSFITKNDPAGGNKIKAVGSKP 286
DB 152 HKFNVLASQPADFD-----RIKGLN-----MQNLLTAHPD-----VOAV---- 187
QY 287 ASIFKGLAPNDGMAEQATKLEGFDTQKIFVTRQDYNDKAKFTFKDQDNMTIYK-P 345
DB 187 -----FAQNDMAALGALRALQTAG--RADVNVVGFDTPDGEEKAVKDGKLAATIAQLP 237
QY 346 DKVLGKVAVEVLRLVLIKAKNKA 367
DB 238 DQITAKGVETADKVLKGEVQA 259
RESULT 9
S56453
hypothetical 32.1k protein (ppa-fbp intergenic region) precursor (o318) - Escherichia
C;Species: Escherichia coli
C;Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 14-Nov-1997
C;Accession: S56453; F85234
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from
A;Reference number: S56314; MUID:95334362
A;Accession: S56453
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-318 <BLAT>
A;Cross-references: EMBL:U14003; NID:g1263172; PID:g537069
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: F65234
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-318 <BLAT>
A;Cross-references: GB:AE000494; GB:U00096; NID:g1790670; PID:g1790674; UWGP:b4227
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: ytfQ
Query Match 5.5%; Score 118; DB 2; Length 318;
Best Local Similarity 20.4%; Pred. No. 0.6;
Matches 78; Conservative 48; Mismatches 133; Indels 124; Gaps 13;
QY 7 KFLYSALYATSLASIIAFVAGCGQTESGSDSKPQAEITLKHVSNDSRIALTD 66
DB 3 KRLIIVSAVSA-AMSSMALAAPLTVGFVSQVSGESG----- 37
QY 67 NPRWISAOKDIISYVDETEAATSTITKNQDQANNWL--TOANLSPAPKGFIIAPENGSGV 125
DB 37 ---WRAETNVAKSEAKRGITLKIADGQKQENQIKAVRSFVAQGVDAIFAPVATGW 93
QY 126 GTAVNTIADKGIPIVAYDRILITGSDRY-----DWYVSFDNEK-----VG 164

Db 94 EPVLKEAKDAEIPVFLDRSDIVKDSLYMTVTADNILEGKLLIGDLWLVKEVNGKPCNVV 153

QY 165 ELOGLSAAGLLGKEDGAFDSI-----DOMNEYLSKSHMPQETISFYTIAGSQDDN 214

Db 154 ELQGTGASVAIDRRKGFABAKNAPNIKIIRSQSGDFTRSKGKVEWESI-----KAEN 208

QY 215 NSQ-----YFVN-----GAMVKELMKNSQNKIIDLSPEGENAVYVPGWNYGTAGORIQ 264

Db 209 NGKNICMYAHNDMDVIGAIQAIKEAGLKFKGKIDLTGSDIGVFDIY----- 255

QY 265 SFLUTINKDPAGGNKIKAVGSKPASIFKFLAPN-DGMAEQAITKLKLEGFDTKIFVTRQ 323

Db 255 -----KAMMDGEANASV-----ELTPNWAGPFALEKFKKDGTMPEKLTITKS 298

QY 324 DYNDKAKTFIKDGDONMTIYKPD 346

Db 299 -----TLVLPD 304

RESULT 10

S54395

permease-like protein - Streptococcus pneumoniae (fragment)

C:Species: Streptococcus pneumoniae

C:Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Sep-1997

C:Accession: S54395

R:Pearce, B.J.; Naughton, A.M.; Masure, H.R.

Mol. Microbiol. 12, 881-892, 1994

A:Title: Peptide permeases modulate transformation in Streptococcus pneumoniae.

A:Reference number: S54395; MUID:95020610

A:Accession: S54395

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-643 <PEA>

A:CROSS-references: EMBL:L20556; NID:g516639; PID:g516640

C:Genetics:

A:Gene: p1pa

Query Match 5.5%; Score 118; DB 2; Length 643;

Best Local Similarity 21.1%; Pred. No. 1.7;

Matches 90; Conservative 60; Mismatches 163; Indels 114; Gaps 20;

QY 27 VAAGCGTSGTSKSDSPQAEITLKHKVSNDISRIALTDPNPRWISAQKDIISYVDTEA 86

Db 2 VLAAC-----SGSGSSAKGE-KTFY-----IYETDPDN-----LNLATATA 38

QY 87 ATSTITKN-----QDAQNNWLTQQANLSPAPKGFIIAPENGSGVGTAVNTIADKGIPI 139

Db 39 ATANITSNVVDGLLENDRYGNFVPSMA-----EDWS-----VSKDGL-- 76

QY 140 VAYDRLITGSKDYWYVDFNFKVGELOGLSLAAGLLGKEDGAFDSIDQMEYLK-----SH 196

Db 76 -TYTITRKDAK--WYTS-EGEEVAAVKAODFTVGLAYADKSDALYPQESIKGLDAY 131

QY 197 MPQETISFYIIA-GSQDDNNSQFYNGAMVKELMKNSQNKIIDLSPEGENAVYVPGWN 255

Db 132 VKGEIKDFSOVGKALDEQTVQTLN-----KPESEFWSNXTTMGVLPAPVNEEFLNSKGD 186

QY 256 YGTAGORIQSFLITINKDPA-----GGNKIKAVGSKPASIEFKFLAPNDGMAEQ-----AITEK 307

Db 187 FAKA-----TPSSLLYNGPYLLKSIYTKSVSEF-----AKNPNTWMDNDNHIDK 231

QY 308 LKLEGFDTKIFVTRQDYNDAKT-----FIKDGONMTIYKPKDKVLGKVAV 354

Db 232 VKLSFWDQDTSKPAENFKDGLSLAARLYPTSASFALKSMKDNIVYTQODSITILVGT 291

QY 355 EVLRLVIAKNKASRSVENELRAKLPN-----ISFYKIDNOTYKVQ-----GKNINTI 402

Db 292 NIDRSQYKTSKTSDEQKASTKKALLNKDFRQALAFGDFRTAYASQLNGQTGASKILRLN 351

QY 403 LVSPVIV 409

Db 352 FVPPTFV 358

RESULT 11

JGECR

periplasmic ribose-binding protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 13-Feb-1998

C:Accession: A03425; H65178; S72655

R:Groarke, J.M.; Mahoney, W.C.; Hope, J.N.; Furlong, C.E.; Robb, F.T.; Zalkin, H.; He

J. Biol. Chem. 258, 12952-12956, 1983

A:Title: The amino acid sequence of D-ribose-binding protein from Escherichia coli K1

A:Reference number: A03425; MUID:84032513

A:Accession: A03425

A:Molecule type: DNA

A:Residues: 1-296 <GRO>

A:Experimental source: strain K12 DG50-3

A>Note: most of the primary structure was confirmed by protein sequencing

R:Mcbray, S.L.; Cole, L.B.

J. Mol. Biol. 225, 155-175, 1992

A:Title: 1.7 angstroms x-ray structure of the periplasmic ribose receptor from escher

A:Reference number: A44680

A:Contents: annotation: X-ray crystallography, 1.7 angstroms

R:Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: H65178

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-296 <BLAT>

A:CROSS-references: GB:AE000452; GB:U00096; NID:g1790189; PID:g1790192; UWGP:b3751

A:Experimental source: strain K-12, substrain MG1655

R:Gonzalez-Gil, G.; Bringmann, P.; Kahmann, R.

Mol. Microbiol. 22, 21-29, 1996

A:Title: FIS is a regulator of metabolism in Escherichia coli.

A:Reference number: S72651; MUID:97055418

A:Accession: S72655

A:Molecule type: protein

A:Residues: 26-39 <GON>

C:Comment: This periplasmic binding protein is involved in the high-affinity D-ribose

C:Genetics:

A:Gene: rbsB

C:Superfamily: lac repressor

C:Keywords: chemotaxis; periplasmic space; sugar transport

F:1-25/Domain: signal sequence status predicted <SIG>

F:26-296/Product: D-ribose-binding protein #status experimental <MAT>

F:26-128,261-289/Domain: 1 #status experimental <TOP>

F:129-260,290-296/Domain: 2 #status experimental <BOT>

F:38,114,115,166,215,240,260/Binding site: beta-D-ribofuranose (Asn, Asp, Arg, A

Query Match 5.5%; Score 117.5; DB 1; Length 296;

Best Local Similarity 21.2%; Pred. No. 0.59;

Matches 76; Conservative 55; Mismatches 138; Indels 89; Gaps 14;

QY 20 LASIIAFVAGCGQTSGTSKSDSPQAEITLKHKVSNDISRIALTDPNPRWISAQKDIIS 79

Db 6 LATVLSAVAL-----SATVSANAMAKDITLVVSTLNNPFFVSLK----- 46

QY 80 YVDEEATATSTITKN---QDAQNNWLTQQANLSP-----APKGFIIAPENGSGVGTAVNTI 132

Db 46 --DGAQKEADKGLYNLVLDSSQNNPAKELANVQDLTVRGTKILLINPTSDAVGNVAKMA 103

QY 133 AKGIPDIYADRLITGSKDYWYVDFNFKVGELOGL--LSLAAGLLGKEDGAFDSIDQMN 190

Db 104 NOANIPVITLDROATKGEVVS-HIASDNYLGGKIAGDYIAKAG-----EGA----- 150

QY 191 EYLSKSHMPQETISFYTIAGSQDDNNSQFYNGAMVKELMKNSQNKIIDLSPEGENAVY 250

Db 150 -----KVIELQGIAGTSARERGERGQQAVAAHKFNVLASQPADFD-----R 191

QY 251 VPGWNYGTAGORIQSFLITINKDPAGGNKIKAVGSKPASIEFKFLAPNDGMAEQAITK 310

Db 192 IKGLNV-----MQLLTAHPD-----VQAV-----FAQDEMALGALRALQT 228
Qy 311 EGFDTQKIFVTRQDYNDAKAKFIKDGQDNMTYK-PDKVLGKVAQVLRVLVLIKKNKA 367
Db 229 AG--KSDVMVVGFDGTPDGERKAVNDGKLAATLAQLPDQIGAKGVETADKVLKGEKVOA 284
RESULT 12
A64073
D-ribosome-binding protein - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 17-Oct-1997
C:Accession: A64073
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: A64073
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-292 <TIGR>
A:Cross-references: GB:U32732; GB:L42023; NID:g1573480; PID:g1573485; TIGR:HI0504
C:Superfamily: lac repressor
C:Keywords: chemotaxis; periplasmic space; sugar transport
F:36,112,113,164,213,237,257/Binding site: beta-D-ribofuranose (Asn, Asp, Arg, Arg, Asn,
Query Match 5.4%; Score 116.5; DB 2; Length 292;
Best Local Similarity 19.3%; Pred. No. 0.68;
Matches 74; Conservative 57; Mismatches 128; Indels 125; Gaps 15;
Qy 20 LASITFAVAGCGQTSGSTSDSKPQAE TLKHKVNDSIRIALTPDNPRIWISAKQDIIS 79
Db 4 LTALTSAVLLGLAVSSAS-----AQDTIALAVSLDNPFFVTLKDGAK 48
Qy 80 YDETEATSTITKNQDAQNWLTOQANLSP-----APKGFIIAPENGSGVTA VNTIADK 135
Db 49 KADELGYKLVL-----DSQNDPAKELANIEDLTVRGAKILLINTPASEAVGNAVAIANRK 104
Qy 136 GIPIVAYDRILITGSDKYDWMYVDFNEKVGELQGLSLAAGLLGKEDGAFDSIDOMNEYLKS 195
Db 105 HIPVITLDR-----GAAGK-----NVVS 122
Qy 196 HMQPTISFYTAGSQDDNNSQYFNGAMKVLKELMKNKSONKIIDLSPGEGNAVYVPGWN 255
Db 123 HIASDNIAGGRKMGD-----FIAQKLGDNA--KVIQL--EG-----IAGTS 159
Qy 256 YGTAGQRIQSFLTINKDPAGGNKIKAVGSKPAS I-----FKGFLAPN 297
Db 160 --AAREREGF---KQALDAHKFNVLASQPADPRTKGLNVTENLLASKGDGVQVIFAQN 213
Qy 298 DGMAEQAITKLKLEGFTQKIFVTRQDYNDAKAKFIKDGQDNMTYKPKDVLGKVAQVLR 357
Db 214 DEMALGALRAVKA--NKKVLIVGFDGDDGKAVKSKMAATTAQPELIGSLGV--- 268
Qy 338 RVLIAKKKNKASSEVENELKALP 381
Db 268 ----VTADKILKGE---KVEAKIP 284
RESULT 13
JGECG
D-galactose-binding periplasmic protein precursor - Escherichia coli
N:Alternate names: mglB protein
C:Species: Escherichia coli
C:Date: 02-Apr-1982 #sequence_revision 18-Nov-1994 #text_change 14-Nov-1997
C:Accession: A37277; A92319; A94599; A30384; A32653; E64983; A03427; Q00294
R:Hogg, R.W.; Voelker, C.; Von Carlowitz, I.
Mol. Gen. Genet. 229, 453-459, 1991
A:Title: Nucleotide sequence and analysis of the mgl operon of Escherichia coli K12.
A:Reference number: A37277; MUID:92049246
A:Accession: A37277
A:Molecule type: DNA
A:Residues: 1-332 <HOG>
A:Cross-references: GB:M59444; NID:gl46852; PID:gl46853
A:Experimental source: strain K12
R:Mahoney, W.C.; Hogg, R.W.; Hermodson, M.A.
J. Biol. Chem. 256, 4350-4356, 1981
A:Title: The amino acid sequence of the D-galactose-binding protein from Escherichia
A:Reference number: A92319; MUID:81168234
A:Accession: A92319
A:Molecule type: protein
A:Residues: 24-178, 'KE', 181-332 <MAL>
A:Experimental source: strain B/r
A:Note: this sequence revised in reference A94599
R:Mahoney, W.C.; Hogg, R.W.; Hermodson, M.A.
submitted to the Atlas, November 1982
A:Reference number: A94599
A:Contents: revisions
A:Accession: A94599
A:Molecule type: protein
A:Residues: 24-332 <MA2>
R:Scholle, A.; Vreemann, J.; Blank, V.; Nold, A.; Boos, W.; Manson, M.D.
Mol. Gen. Genet. 208, 247-253, 1987
A:Title: Sequence of the mglB gene from Escherichia coli K12: comparison of wild-type
A:Reference number: A30384; MUID:87286407
A:Accession: A30384
A:Molecule type: DNA
A:Residues: 1-15 <SCH>
R:Scripture, J.B.; Hogg, R.W.
J. Biol. Chem. 258, 10853-10855, 1983
A:Title: The nucleotide sequences defining the signal peptides of the galactose-bind
A:Reference number: A32653; MUID:83291030
A:Accession: A32653
A:Molecule type: DNA
A:Residues: 1-15 <SCR>
R:Vyas, N.K.; Vyas, M.N.; Quijcho, F.A.
Proc. Natl. Acad. Sci. U.S.A. 80, 1792-1796, 1983
A:Title: The 3 angstrom resolution structure of a D-galactose-binding protein for tra
A:Reference number: A93942; MUID:83169767
A:Contents: annotation; X-ray crystallography, 3.0 angstroms
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E64983
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-332 <BLAT>
A:Cross-references: GB:AE000304; GB:U00096; NID:g1788470; PID:g1788473; UWGP:b2150
A:Experimental source: strain K-12, substrain MGI655
C:Genetics:
A:Gene: mglB; mglM
A:Map position: 46 min
C:Superfamily: D-galactose-binding protein
C:Keywords: calcium binding; chemotaxis; periplasmic space; sugar transport
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-332/Product: D-galactose-binding protein #status experimental <MAT>
F:37,114,175,177,181,234,259,279/Binding site: galactose (Asp, Asn, His, Asp, Arg, As
F:157,159,161,163,165,228/Binding site: calcium (Asp, Asn, Asp, Gln, Gln, Glu) #statu

Db 36 DDN--FMSVVRKAIE--ODAKAAPDVOLLMNDSDNDOSKONDOIDVLLAKGKALAINLV 91
QY 121 NGSGVGTAVNTIADKGIPIVAYDRLIT--GSDKYD--WTVSFDNEKVGELGSLAAGLL 176
Db 92 DPAAGTVEKARGQNVVFFNKESRKALDSYKRAYIVGTDSKESGIQG-----DLI 146
QY 177 KREGAFDSIDQNEYLKSHMPQETISFYTIAG--SQDDNNSOYFYNGAMKVLKEL----- 231
Db 147 AKHWAANQWD-LNK-----DGQIQFVLLKGEPGHPDAEARTTY-----VIKELNDKG 193
QY 231 MKNSONKIIDLSPEGENAVYVPGWNYGTAGQRIQSFLTINKDPAGNKIKAVGSRPASIF 290
Db 194 IKTEQIQQL-----DTAM-----WDTAQAKDKMDAWLS-----GPNANKIEVV----- 231
QY 291 KGFAPNDGMAEQAIKLEGFDTQKIFVTRQDYNDKAKTFIKDQGNMTIYKPKDKVLG 350
Db 231 ---IANDAMAMGAVEALKAHNKSSIPVFGV--DALPEALALVKSALAGTIVLNDANNOA 285
QY 351 KVAVEVLRVLIKKNKASRS--EVENELKAKLPNISFKYDN 389
Db 286 KATFDLAKNLADGKAADGTNNKIDNKV--VRVPYGVGDKDN 325

RESULT 14

B70126
surface-located membrane protein 1 (lmp1) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
C:Accession: B70126
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:198065943
A:Accession: B70126
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1119 <KLE>
A:C:References: GB:AE001131; GB:AE000783; NID:g2688098; PID:g2688100; TIGR:BB0210
A:Experimental source: strain B31
C:Superfamily: tetratricopeptide repeat homology
F:742-774/Domain: tetratricopeptide repeat homology #status atypical <TT1>
F:775-808/Domain: tetratricopeptide repeat homology <TT2>
F:809-842/Domain: tetratricopeptide repeat homology <TT3>
F:843-876/Domain: tetratricopeptide repeat homology <TT4>
F:911-943/Domain: tetratricopeptide repeat homology <TT5>
F:944-977/Domain: tetratricopeptide repeat homology <TT6>
F:979-1010/Domain: tetratricopeptide repeat homology #status atypical <TT7>
F:1011-1044/Domain: tetratricopeptide repeat homology <TT8>

Query Match 5.3%; Score 113.5; DB 2; Length 1119;
Best Local Similarity 20.2%; Pred. No. 7.3;
Matches 101; Conservative 63; Mismatches 156; Indels 179; Gaps 22;

QY 47 ETLKHVV-----SNDSIRIALTD-----PDNPRWISAOKDI-----ISVDETEAA 87
Db 650 EDUKSVHSIKPIDLENTKSRQAIAKDLNEFLKNPNDAQAASKTLAQANKIOHLEDKSK 709
QY 88 TSTITNQDAQNNWLTQQA-----NLSAPKGFIIAPENGSGVGTAVN----- 131
Db 710 VHSI-KPIDLENTKSRQAIAKDLNEFLKNPNDAQAASKTLAQAYENNGLLKAENAYEKI 768
QY 131 -----TIADKGIPIVAYDRLITG 148
Db 769 IKLTNTQEDHYKLGIIIRFKKKYEHSGIESFDQTKIDPKHKHALHNKGIALMMLNKNKA 828
QY 149 SKDYDWTVSFD-----NEKVGELQ-----GLSLAAGLLGKE 179
Db 829 IESFEKAIQIDKNYGTAYYQKGTAEKNGDMQQAFAFKNAYNLDKNPNYALKAGIVSN 888

QY 180 DGAFDSIDQNEYL-----KSHMPQETISFYTIAGSQDDNN----- 216
Db 889 LGNF---KQSEYLNFFNANAKPNE-IAIYNLSIAKFNENKLESETINKAIDLNPEK 944
QY 216 SOYFYNGAMKVLKELKNSQNKI-----IDLSPGENAVYVPGWNYGTAGQRIQSFLT 269
Db 945 SEYLYLKASINLKK--ENYQNAISLYSLVIEKNPENTSAIYNLAKAYEKSGNKSQAISTL 1002
QY 270 NKDPAGGNK--IKAVG-----SKPASIFKGLFAPNDGMAEQAIKLEGFDTQKI 318
Db 1003 EKIIKNNKLNALNNLIYKKEKNYOKAIEF-----EKAIINSIDIEA--KNL 1049
QY 319 FVTRQDYNDKAKTFIKDQGNMTIYKPKDKVLGKVAVEVLRVL-IAKKNKASRSSEYENELK 377
Db 1050 ATTLEINDNTRA--KDLLREYTKLKPNN-----PEALHALGIYEYNNNDQTLRELI 1101
QY 378 AKLPNISFKYDNTYKVOG 396
Db 1102 KKFPN--YKKNENIKKIIG 1118

RESULT 15

S29390
galactose-binding protein - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 01-Dec-1995
C:Accession: S29390
R:Benner-Luger, D.; Boos, W.
Mol. Gen. Genet. 214, 579-587, 1988
A:Title: The mglB sequence of Salmonella typhimurium L72; promoter analysis by gene f
A:Reference number: S29389; MUID:89112167
A:Accession: S29390
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <BEN>
C:Superfamily: D-galactose-binding protein

Query Match 5.3%; Score 113; DB 2; Length 332;
Best Local Similarity 20.3%; Pred. No. 1.4;
Matches 84; Conservative 61; Mismatches 152; Indels 116; Gaps 19;

QY 5 LRKKELYSSAIYATSLASTIAFVAAGCGQTSGST-----SDSKPQ 45
Db 1 MNKKVLTLSAV-----MASLLFGAHAHAADTRIGVTIYKYDDNFMVVRKAIEKDGKSAPD 56
QY 46 AETLKHVSNISIRIALTDPDNPRWISAOKDIISVDETEATSTITKNQDAQNWLTTQ 105
Db 57 VOLLMNDSDNQ-----SKNDQIDVL----- 79
QY 106 ANLSPAPKGFIIAPENGSGVGTAVNTIADKGIPIVAYDRLIT--GSDKYD--WYVSFDE 161
Db 79 --LAIGVKAALNVLDPAAAGTVIEKARGQNVVFFNKESRKALDSYKRAYIVGTGTSK 136
QY 162 KYGELQGLSLAAGLLKEDGAFDSIDQNEYLKSHMPQETISFYTIAG--SQDDNNSOYF 219
Db 137 ESGVIQG-----DLIAKHQANQGW-LNKDGK-----IQYVLLKGEPGAPDAEARTT 183
QY 220 YNGAMKVLKELM-KNSQNKIIDLSPEGENAVYVPGWNYGTAGQRIQSFLTINKDPAGGNK 278
Db 184 Y-----VVKELNDKGIQTEQLAL-----DTAM-----WDTAQAKDKMDAWLS-----GPNANK 226
QY 279 IKAVGSKPASIFKGFAPNDGMAEQAIKLEGFDTQKIFVTRQDYNDKAKTFIKDQGD 338
Db 227 IEVV-----IANDAMAMGAVEALKAHNKSSIPVFGV--DALPEALALVKSAM 273
QY 339 NMTIYKPKDKVLGKVAVEVLRVLIKKNKASRS--EVENELKAKLPNISFKYDN 389
Db 274 ACTVLNDANNQAKATFDLAKNLAECKGAADGTSWKIENKI--VRVPYGVGDKDN 325

Search completed: November 13, 1999, 12:07:50
Job time: 2062 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 09:07:10 ; Search time 51.07 Seconds
(without alignments)
231.925 Million cell updates/sec

Title: US-08-913-430-2
Perfect score: 2143
Sequence: 1 MKMKELLYSAIYATSL.....NTILSPVIVTKANDNPDA 419

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2121	99.0	416	1 P46_MYCHY	P46192 mycoplasma
2	335	15.6	354	1 CVE2_AGRU	P54082 agrobacteri
3	326	15.2	332	1 XYLE_HAEIN	P45047 haemophilus
4	323	15.1	354	1 CVE1_AGRU	P25548 agrobacteri
5	305.5	14.3	330	1 XYLE_ECOLI	P37387 escherichia
6	305	14.2	357	1 CHVE_AZOB	P54083 azospirillum
7	123	5.7	332	1 DGAL_CITFR	P23823 citrobacter
8	118.5	5.5	660	1 ALIA_STRPN	P35592 streptococc
9	118	5.5	318	1 YTFQ_ECOLI	P39325 escherichia
10	117.5	5.5	296	1 RBSB_ECOLI	P20925 escherichia
11	116.5	5.4	292	1 RBSB_HAEIN	P44737 haemophilus
12	116	5.4	332	1 DGAL_ECOLI	P02927 escherichia
13	113	5.3	332	1 DGAL_SALT	P32905 salmonella
14	112.5	5.2	3135	1 S230_PLAFO	Q08372 plasmodium
15	112	5.2	296	1 RBSB_YEAST	P02926 salmonella
16	111.5	5.2	1451	1 SPT6_YEAST	P23615 saccharomyc
17	111	5.2	1104	1 COLA_CLOPE	P43153 clostridium
18	109.5	5.1	993	1 EPB2_MOUSE	P54763 mus musculu
19	108.5	5.1	986	1 EPB2_HUMAN	P29323 homo sapien
20	108	5.0	1957	1 YD86_SCHPO	Q10411 schizosacch
21	107	5.0	1182	1 CGA2_HELPY	P55746 helicobacte
22	107	5.0	331	1 DGAL_HAEIN	P44883 haemophilus
23	106.5	5.0	315	1 FHUD_BACSU	P37580 bacillus su
24	106.5	5.0	802	1 SUSL_MAIZE	P04712 zea mays (m
25	106.5	5.0	1225	1 Y309_MTCGE	P47551 mycoplasma
26	105.5	4.9	2491	1 TALA_DICDI	P54633 dictyosteli
27	104.5	4.9	212	1 RBSB_BACSU	P36949 bacillus su
28	104	4.9	460	1 EFLA_SORMA	Q09069 sordaria ma
29	104	4.9	1518	1 KKK1_YEAST	P34244 saccharomyc
30	103.5	4.8	460	1 EFLA_TREIR	P34825 trichoderma
31	103.5	4.8	1407	1 T2D2_YEAST	P23255 saccharomyc
32	103	4.8	1337	1 DEXT_STRDO	P39653 streptococc
33	103	4.8	456	1 GCAD_BACSU	P14192 bacillus su
34	103	4.8	591	1 Y040_MYCGE	P47286 mycoplasma
35	103	4.8	657	1 Y040_MYCPN	P75062 mycoplasma
36	102	4.8	641	1 SAC2_YEAST	P39904 saccharomyc
37	101	4.7	1155	1 IF2P_METJA	Q57710 methanococc
38	101	4.7	762	1 SLAP_ACEKI	P22258 acetogenium
39	100	4.7	335	1 MALR_CLOBU	O05103 clostridium
40	100	4.7	1176	1 SLAP_BACSH	P38537 bacillus sp
41	100	4.7	660	1 VG06_BPT4	P19060 bacterioph
42	100	4.7	571	1 XINC_PSEFL	P23031 pseudomonas
43	99.5	4.6	931	1 DAPI_YEAST	P33894 saccharomyc

44	99.5	4.6	1430	1	GTFD_STRMU	P49331 streptococ
45	99	4.6	877	1	EPA5_MOUSE	Q60629 mus musculu
ALIGNMENTS						
RESULT 1						
P46_MYCHY	STANDARD;	PRT;	416 AA.			
AC P46192:						
DT 01-NOV-1995 (REL. 32, CREATED)						
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)						
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)						
DE 46 KD SURFACE ANTIGEN PRECURSOR (P46).						
OS MYCOPLASMA HYOPNEUMONIAE.						
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLICUTES;						
OC MYCOPLASMATACEAE; MYCOPLASMA.						
RN [1]						
RP SEQUENCE FROM N.A., AND SEQUENCE OF 67-74 AND 338-358.						
RC STRAIN-ATCC 25934 / J;						
RX MEDLINE; 95204368.						
RA FUTO S., SETO Y., MITSUSE S., MORI Y., SUZUKI T., KAWAI K.;						
RT "Molecular cloning of a 46-kilodalton surface antigen (P46) gene from						
RT Mycoplasma hyopneumoniae: direct evidence of CGG codon usage for						
RT arginine.";						
RL J. BACTERIOL. 177:1915-1917(1995).						
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID ANCHOR						
CC (POTENTIAL).						
CC						
CC This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -						
CC the European Bioinformatics Institute. There are no restrictions on its						
CC use by non-profit institutions as long as its content is in no way						
CC modified and this statement is not removed. Usage by and for commercial						
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/						
CC or send an email to license@isb-sib.ch).						
CC						
CC EMBL; D16682; G1220114; ALT INIT.						
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.						
DR PFAM; PF00532; Peripla_BP_like; 1.						
KW ANTIGEN; MEMBRANE; LIPOPROTEIN; SIGNAL.						
FT SIGNAL 1 27 POTENTIAL.						
FT CHAIN 28 416 46 KD SURFACE ANTIGEN.						
FT LIPID 28 28 N-ACYL DIGLYCERIDE (POTENTIAL).						
SQ SEQUENCE 416 AA; 45401 MW; 3EA04FED CRC32;						
Query Match 99.0%; Score 2121; DB 1; Length 416;						
Best Local Similarity 99.8%; Pred. No. 3.3e-127;						
Matches 415; Conservative 0; Mismatches 1; Indels 0; Gaps 0;						
QY 4	MURKELYSSAIYATSLAIIAFVAGCGGTSGTSDSKPQAEIHKHKVNSDSIRIALT	63				
DB 1	MURKELYSSAIYATSLAIIAFVAGCGGTSGTSDSKPQAEIHKHKVNSDSIRIALT	60				
QY 64	DPDNPWISAQKDIIISYVDETAATSTITKNQDANNWLTQQANLSPAPKGFIIAPENG	123				
DB 61	DPDNPWISAQKDIIISYVDETAATSTITKNQDANNWLTQQANLSPAPKGFIIAPENG	120				
QY 124	GVGTAVNTIADKGIPIVADRLITGSDKYDWTYVDFNEKVGELQGLSLAAGLLKEDGAF	183				
DB 121	GVGTAVNTIADKGIPIVADRLITGSDKYDWTYVDFNEKVGELQGLSLAAGLLKEDGAF	180				
QY 184	DSIDOMNEYLKSHMPQETISFYTIAGSQDDNNQSYFYNGAMKVLKELMNSQNKIIDLSP	243				
DB 181	DSIDOMNEYLKSHMPQETISFYTIAGSQDDNNQSYFYNGAMKVLKELMNSQNKIIDLSP	240				
QY 244	EGENAVYVPGWNYGTAGRIQSFLLINKDPAGNKKIKAVGSKPASTFKGFLAPNDGMAEQ	303				
DB 241	EGENAVYVPGWNYGTAGRIQSFLLINKDPAGNKKIKAVGSKPASTFKGFLAPNDGMAEQ	300				
QY 304	AITKLKLEGFDTKIFVTRODYNDRKAKTFIKDGDQNMNTIYKPKVLGLGVAVEVLRVLI	363				

```

Db 301 AITKLEGEFTQIFVTDYNDKAKTFIKDQNMNTIYKPKDVLGVKVAEVLRLVIAK 360
QY 364 KNAKSRSEVENELAKLPNISFKYDQNTYKVOGKNINTILVSPVIVTKANVDNPA 419
Db 361 KNAKSRSEVENELAKLPNISFKYDQNTYKVOGKNINTILVSPVIVTKANVDNPA 416

RESULT 2
CUE2_AGRU
ID CUE2_AGRU STANDARD; PRT; 354 AA.
AC P34082;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MULTIPLE SUGAR-BINDING PERIPLASMIC RECEPTOR CHVE PRECURSOR.
GN CHVE.
OS AGROBACTERIUM TUMEFACIENS.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D10B/87;
RA BELANGER C., LOUBENS I., NESTER E.W., DION P.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: REQUIRED FOR EFFECTIVE TRANSCRIPTIONAL INDUCTION OF THE
CC VIR GENES BY MONOSACCHARIDES IN RESPONSE TO PLANT SIGNALS AND FOR
CC NORMAL GROWTH AND CHEMOTAXIS TOWARDS CERTAIN SUGARS. FUNCTION AS A
CC PERIPLASMIC MULTIPLE SUGAR-BINDING RECEPTOR PROTEIN. IT DOES NOT
CC INTERACT WITH A TRANSPORT SYSTEM.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC RECEPTOR FAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X97456; E237609; -
CC DR PFAM; PF00532; Peripla_BP_like; 1.
CC KW SIGNAL; CHEMOTAXIS; PERIPLASMIC; CROWN GALL TUMOR.
CC FT SIGNAL 1 25 POTENTIAL.
CC FT CHAIN 26 354 MULTIPLE SUGAR-BINDING PERIPLASMIC
CC SQ SEQUENCE 354 AA; 37729 MW; 7679/A2F CRC32;

Query Match 15.6%; Score 335; DB 1; Length 354;
Best Local Similarity 27.4%; Pred. No. 1.8e-14;
Matches 110; Conservative 68; Mismatches 155; Indels 68; Gaps 12;

QY 19 SLASIIAFVAGCGQTSGSTSDSKPOAETLKHVKVNSDRIALTDPNRPNWISAQKDII 78
Db 3 SIILSTAAAIQVAMFVAPAFADK-----GTVGAMPTRKASARWIDGNNIV 50
QY 79 SYVDTEAATSTIKNOAQNNWLTQANL-SPAPKFIITAPENGSGVGTAVNTIADKI 137
Db 51 KQLQAAGYGTDLQYGGDDIPNQ-LSQVENMVTKGDKVLVIAADIGTTLSDVLQKAHDAGI 109
QY 138 PIVAYDRLITGSKDYKDWVSPDNKVGELGSLAAGLKGKDEGDAFSDIQMNEYLKSHM 197
Db 110 KVIAYDLRINSGNVYIATFDNFKVGLQANSIVDG-LGLKDG----- 153
QY 198 PQETISFYTTAGSDDDNNSQYFYNGAMKVLKELMKNQNKIIDLSPGCVNAVYPGWNYG 257
Db 153 -KGFNFIELFGSGPDDNNAFFFYDCAMSVLKPVI-DSGKLWVKSGQQGMDKVTGLRWDP 210
QY 258 TAGORIOSFLTINKDPAGGNKIKVAGSKPASIFKGLFAPNDGMAEQAITKLKLEGGFT-- 316

```

```

Db 211 TAQARMNLLSAYYTD-----HVDV-----LSPYDGLSIGLSLKGKVGIGTGG 256
QY 316 QKI-FVTRDYNDKAKTFIKDQNMNTIYKPKDVLGVKVAEVLRLVIAKNAKSRSEVEN 374
Db 257 QKLPVITGQDSETPSVKSIAGEQHSITIFADTRDLAKVIVDMYNAL----- 303
QY 375 ELKAKLPNISFKYDQNTYKVOGKNINTILVSPVIVTKANVD 415
Db 303 -MEGKTPEVT---DTRKYDNGVKVPSYLLTPVAVDKTNVE 339

RESULT 3
XYLF_HAEIN
ID XYLF_HAEIN STANDARD; PRT; 332 AA.
AC P45047;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE D-XYLOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR.
GN XYLF OR HI1111.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE: 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCCANN K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.";
RL SCIENCE 269:496-512(1995).
CC -1- FUNCTION: INVOLVED IN THE HIGH-AFFINITY D-XYLOSE MEMBRANE
CC TRANSPORT SYSTEM. BINDS WITH HIGH AFFINITY TO XYLOSE
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC RECEPTOR FAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32791; G1574665; -
CC DR TIGR; HI1111; -
CC DR PFAM; PF00532; Peripla_BP_like; 1.
CC KW TRANSPORT; SUGAR TRANSPORT; PERIPLASMIC; SIGNAL.
CC FT SIGNAL 1 23 BY SIMILARITY.
CC FT CHAIN 24 332 D-XYLOSE-BINDING PERIPLASMIC PROTEIN.
CC SQ SEQUENCE 332 AA; 35901 MW; 4B05332D CRC32;

Query Match 15.2%; Score 326; DB 1; Length 332;
Best Local Similarity 28.2%; Pred. No. 6.3e-14;
Matches 108; Conservative 63; Mismatches 118; Indels 94; Gaps 15;

QY 42 SKPOAETLKHVKVNSDRIALTDPNRPNWISAQKQIISYVDETEAATSTI---TKNQD-- 97
Db 19 SSAHSDKLIGLSIDDLRL-----ERW---QKDRDIFVNKAESGAKVFGVQANGDSDS 68
QY 97 AQ-----NNWLTQANLSPAPKFIITAPENGSGVGTAVNTIADKIPIVAYDRLITGSKDY 152

```

```
Db 69 A01SQIENMINKNDV-----LVIPHNCEVLSNVISEAKKGIKVLAYDRLINNAD-L 121
QY 153 DWYVSFNEKVGELQGLSLAAGLLGKEDGAFDSIDOMNEYLAKSHMPOETISFYTIAGSQD 212
Db 122 DFVVSFNEKVGELQGLSLAAGLLGKEDGAFDSIDOMNEYLAKSHMPOETISFYTIAGSQD 212
QY 213 DNNQVFNKVGELQGLSLAAGLLGKEDGAFDSIDOMNEYLAKSHMPOETISFYTIAGSQD 212
Db 158 DNNKLFKRGKQVLPDPLIAGKIKV-----GDQ--WVDSWLAERKALQIMENALTANK- 210
QY 273 PAGNKIKAVGSPASIFKGLFAPNDGMAQATIKLKGFDQKIFVTRDYNDAKTF 332
Db 210 ----NNVDV-----VASNDATAGCAIQALSAQGL-SGVNAISQDADLAIAKR 253
QY 333 IKDQDNMTIYKPKVLGKVAEVLRLVIAKKNKASSEVENELKAKLPNISFYDNQTY 392
Db 254 IVNGSOTMTYKPKVLGKVAEVLRLVIAKKNKASSEVENELKAKLPNISFYDNQTY 392
QY 393 KVQGNKNTILVSPVITKANVD 415
Db 298 ----KNVPAYLLDPIAVDKRNIN 316

RESULT 4
CVEL_AGRU STANDARD; PRT; 354 AA.
AC P25548;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MULTIPLE SUGAR-BINDING PERIPLASMIC RECEPTOR CHVE PRECURSOR.
GN CHVE.
OS AGROBACTERIUM TUMEFACIENS.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C581C1M;
RX MEDLINE: 94075346.
RA SHIMODA N., TOYODA-YAMAMOTO A., AOKI S., MACHIDA Y.;
RT "Genetic evidence for an interaction between the Vira sensor protein
RT and the ChvE sugar-binding protein of Agrobacterium."
RL J. BIOL. CHEM. 268:26552-26558(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C581C1M;
RX MEDLINE: 90202696.
RA HUANG M.-L.W., CANGELOSI G.A., HALPERIN W., NESTER E.W.;
RT "A chromosomal Agrobacterium tumefaciens gene required for effective
RT plant signal transduction."
RL J. BACTERIOL. 172:1814-1822(1990).
CC -!- FUNCTION: REQUIRED FOR EFFECTIVE TRANSCRIPTIONAL INDUCTION OF THE
CC VIR GENES BY MONOSACCHARIDES IN RESPONSE TO PLANT SIGNALS AND FOR
CC NORMAL GROWTH AND CHEMOTAXIS TOWARDS CERTAIN SUGARS. FUNCTION AS A
CC PERIPLASMIC MULTIPLE SUGAR-BINDING RECEPTOR PROTEIN. IT DOES NOT
CC INTERACT WITH A TRANSPORT SYSTEM.
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -!- INDUCTION: BY L-ARABINOSE, D-GALACTOSE, D-FUCOSE IN THE
CC PRESENCE OF GPR (GALACTOSE-BINDING PROTEIN REGULATOR).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC RECEPTOR FAMILY 2.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 212
CC ONWARD AND IS SHORTER (246 AA) DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; D17457; G471095; -
```

```
DR EMBL; M30318; G142283; ALT_FRAME.
DR PFAM; PF00532; Peripla_BP_like; 1.
FT SIGNAL; CHEMOTAXIS; PERIPLASMIC; CROWN GALL TUMOR.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 354 MULTIPLE SUGAR-BINDING PERIPLASMIC
FT RECEPTOR CHVE.
SQ SEQUENCE 354 AA; 38129 MW; CF61DGA5 CRC32;

Query Match 15.1%; Score 323; DB 1; Length 354;
Best Local Similarity 26.8%; Pred No. 1,1e-13;
Matches 107; Conservative 68; Mismatches 156; Indels 68; Gaps 11;

QY 19 SLASIIAFAAGCGQTESGTSKSPQAEHLKHVSNDISIRIALTDPDNPRWISAQKDII 78
Db 3 SIISLMAACAIGAASFAAPAFQDK-----GSVGIAMPTKSSARWIDDCNNIV 50
QY 79 SYVDETEAATSIITKNQDAONNWLTOQANL-SPAPKGFIIAPENGSGVTAIVTADKGI 137
Db 51 KQLEAGYKTDLOAYADDDIPNQ-LSQIENNVTKGVLVIAIDGTLSLSDVLKQAGEQGI 109
QY 138 PIVAYDRLITGSDKYDVIYSFNEKVGELQGLSLAAGLLGKEDGAFDSIDOMNEYLAKSHM 197
Db 110 KVIAVDRLIRNSGDVSYATFDFNFQVGLQATSI-TDKLGLKDG----- 153
QY 198 PQETISFYTIAGSQDNNQSYFYNGAMKVLKELMKNQNKIIDLSPEGENAVIVPGWNYG 257
Db 153 -KGPFNIELFGGSPDDNNNAFFFDGAMSVLKPI-DSGKLVVKSQGMGMDKVGTLRWDP 210
QY 258 TAGQRTQSPFLTINKDPAGCNKIKAVGSPASIFKGLFAPNDGMAQATIKLKGFDQTK 317
Db 211 TAQARMNLLSAYITDA---KVDV-----LSPYDGLSIGISLSLKGVGYGTKD 256
QY 318 ---IFVTRDYNDAKATFIKDGDNMTIYKPKVLGKVAEVLRLVIAKKNKASRSEVEN 374
Db 257 QPLPVVSGQDAEVPVSVKSIAGEQYSTIFEDRELAKVTVMNVNAV----- 303
QY 375 ELKAKLPNISFYDNQTYKVGQKNINTILVSPVITKAN 413
Db 303 -MEGKEPEVN---DTKTYENGKVVPSYLLKPVAVTKEN 337

RESULT 5
XYLF_ECOLI STANDARD; PRT; 330 AA.
AC P37387;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE D-XYLOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR.
GN XYLF OR XYLT.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-67.
RA SUMIVA M., HENDERSON P.J.F.;
RT "The D-xylose binding protein of Escherichia coli."
RL BIOCHEM. SOC. TRANS. 17:553-554(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE: 94316500.
RA SOFTA H.J., BURLAND V., DANIELS D.L., PLUNKETT G. III, BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL NUCLEIC ACIDS RES. 22:2576-2586(1994).
RN [3]
RP SEQUENCE OF 24-71.
RX MEDLINE: 96117780.
RA SUMIYA M., DAVIS E.O., PACKMAN L.C., MCDONALD T.P., HENDERSON P.J.;
RT "Molecular genetics of a receptor protein for D-xylose, encoded by
RT the gene xylF, in Escherichia coli."
RT
```



```

RL RECEPT. CHANNELS 3:117-128(1995).
RN [4]
RP SEQUENCE OF 24-35.
RX STRAIN=K12 / EMG2.
RA MEDLINE: 97443975.
RL LINK A.J., ROBISON K., CHURCH G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL ELECTROPHORESIS 18:1259-1313(1997).
CC CC -!- FUNCTION: INVOLVED IN THE HIGH-AFFINITY D-XYLOSE MEMBRANE
CC CC TRANSPORT SYSTEM. BINDS WITH HIGH AFFINITY TO XYLOSE.
CC CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC CC -!- E.COLI HAS TWO D-XYLOSE TRANSPORT SYSTEMS THAT ACCUMULATE SUGAR
CC CC AGAINST A CONCENTRATION GRADIENT: THE XYLE SYSTEM WHICH UTILIZES
CC CC THE ELECTROCHEMICAL GRADIENT OF PROTONS AND THAT IS INSENSITIVE TO
CC CC COLD OSMOTIC SHOCK AND THE XYLEF SYSTEM THAT USES A HIGH-ENERGY
CC CC PHOSPHATE COMPOUND AND IS SENSITIVE TO COLD OSMOTIC SHOCK.
CC CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC CC RECEPTOR FAMILY 2.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: U000039; G466704; ..
CC CC EMBL: AE0000434; G1789990; ..
CC CC ECODDBASE: C032.4; 6TH EDITION.
CC CC ECOGENE: EG20252; XLYF.
CC CC PFAM: PF00532; Peripla_BP_like: 1.
CC CC TRANSPORT; SUGAR TRANSPORT; PERIPLASMIC; SIGNAL.
CC CC SIGNAL 1
CC CC 23
CC CC 24 330
CC CC SEQUENCE 330 AA; 335734 MW; 9B49A4F9 CRC32;
CC CC -----

```

ID DGAL_CITFR STANDARD; PRT: 332 AA.
AC P23925;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE D-GALACTOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR (GBP) (D-GALACTOSE/
DE D-GLUCOSE BINDING PROTEIN) (GGBP).
GN MGLB.
OS CITROBACTER FREUNDII.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC CITROBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 8090;
RA GALINDO R.L., DAGGETT GARVIN L., HARDIES S.C.;
RL SUBMITTED (MAY-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE ACTIVE TRANSPORT OF
CC GALACTOSE AND GLUCOSE. IT PLAYS A ROLE IN THE CHEMOTAXIS TOWARDS
CC THE TWO SUGARS BY INTERACTING WITH THE TRG CHEMORECEPTOR.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -1- THE CALCIUM-BINDING SITE IS STRUCTURALLY SIMILAR TO THAT OF
CC EF-HAND PROTEINS, BUT IS IN TWO PARTS, WITH THE LAST CALCIUM
CC LIGAND PROVIDED BY GLU-228.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC RECEPTOR FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59389; G40473; -
DR PIR; S15534;
DR PFAM; PF00532; Peripla_BP_like; 1.
DR HSP; P02927; 1GLG.
KW TRANSPORT; SUGAR TRANSPORT; PERIPLASMIC; CALCIUM-BINDING; CHEMOTAXIS;
FT SIGNAL.
FT CHAIN 1 23 D-GALACTOSE-BINDING PERIPLASMIC PROTEIN.
FT CA_BIND 157 165 FIRST PART OF SITE.
FT SITE 227 228 SECOND PART OF SITE.
FT INTERACT WITH MEMBRANE-BOUND TRG SIGNAL
FT TRANSDUCER.
SQ SEQUENCE 332 AA; 35817 MW; 46508D53 CRC32;

Query Match 5.7%; Score 123; DB 1; Length 332;
Best Local Similarity 21.2%; Pred. No. 0.42;
Matches 85; Conservative 69; Mismatches 155; Indels 92; Gaps 20;

QY 5 LRKFLYSSAIYATSLASIAFVAAGCGQTSESTDSKQPAATLKHVNSDIRALTD 64
DB 1 MNKVLTVLSAV-----MASLFGAAHAADTRIGVTI-----YKDDNFMSVVRK- 46
QY 65 PDNPRWISAKDIISYVDETEAATSTTKNDAQNNLTQGAN-----LSPAPKGFIIAPE 120
DB 46 -----AIEKD-----AKAPDVOLLMNDSONDSQKNDQIDVLKAGVKALAINLV 91
QY 121 NGSGVGFVANTIAKGIPIVAYDLRIT--GSDRYD--WYYSFDNEKVGELQGLSLAAGLL 176
DB 92 DPAAAGVIEKARQNPVPIFFENKPSRKALDSYKAYVYGTDSKESGIIQG-----DLI 146
QY 177 GKEDGAFSDIDQNEYLKSHMPQETISFYIAG--SODDNNQSYFYNGAMKVLKEL----- 231
DB 147 AKHWAANPNWD-LNKDGK-----IOFVLLKGPCHPDAEARTY-----VIKELNDKG 193
QY 231 MKNSONKIIDLSPGENAVYVPGWNYGTAGORIOSELTINKDPAGGNKIKAVGSKPASIF 290
DB 194 IKTEQLQL-----DTAM-----WDTAQAKDKMDAAMS-----GPNANKIEVV----- 231

QY 291 KGFLAPNDGMAEQAITKLEGFDTQKIFVTRODYNDKAKTFIKDQDNMTYKPKDKVLG 350
DB 231 ---IANNDAWAGAVEALKAHNKTSPVFGV--DALPEALALVKSGAMAGTVLNDANNQA 285
QY 351 KVAVEVLRLVLIKKKNKASRS--EVENELKAKLPNLSFKYDN 389
DB 286 KATFDLAKNLADGKAADGTNWKIENKI-VRVFYVGVGVDKDN 325

RESULT 8
ALIA_STRPN STANDARD; PRT: 660 AA.
ID AC P35592; Q54782; O54620; O52228;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE OLIGOPEPTIDE-BINDING PROTEIN ALIA PRECURSOR (EXPORTED PROTEIN 1).
GN ALIA OR EXP1 OR PUPA.
OS STREPTOCOCCUS PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R800;
RX MEDLINE; 94328326.
RA ALLOING G., DE PHILIP P., CLAVERYS J.-P.;
RT "Three highly homologous membrane-bound lipoproteins participate in
RT oligopeptide transport by the Ami system of the gram-positive
RT Streptococcus pneumoniae.";
RL J. MOL. BIOL. 241:44-58(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NCTC 11906, SP-496, SP-VA92 AND SP-VA96;
RX MEDLINE; 98125733.
RA COFFEY T.J., ENRIGHT M.C., DANIELS M., MORONA J.K., MORONA R.,
RA HRYNIEWICZ W., PATON J.C., SPRATT B.G.;
RT "Recombinational exchanges at the capsular
RT biosynthetic locus lead to frequent serotype changes among natural
RT isolates of Streptococcus pneumoniae.";
RL MOL. MICROBIOL. 27:73-83(1998).
RN [3]
RP SEQUENCE OF 18-660 FROM N.A.
RC STRAIN-R6X;
RX MEDLINE; 95020610.
RA PEARCE B., NAUGHTON A.M., MASURE H.R.;
RT "peptide permeases modulate transformation in Streptococcus
RT pneumoniae.";
RL MOL. MICROBIOL. 12:881-892(1994).
RN [4]
RP SEQUENCE OF 347-509 FROM N.A.
RC STRAIN-R6X;
RX MEDLINE; 95020625.
RA PEARCE B.J., YIN Y.B., MASURE H.R.;
RT "Genetic identification of exported proteins in Streptococcus
RT pneumoniae.";
RL MOL. MICROBIOL. 9:1037-1050(1993).
CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC FOR OLIGOPEPTIDES; PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z35135; G511063; -
DR EMBL; AF030359; G2804687; -

```
DR EMBL: AF030360; G2804693; -
DR EMBL: AF030361; G2804699; -
DR EMBL: AF030364; G2804719; -
DR EMBL: L20556; G516640; -
DR PROSITE; PS00013; PROKARYOTIC_LIPOPROTEIN; 1.
DR PFAM; PF00496; SBP_Dac_5; 1.
KW PEPTIDE TRANSPORT; TRANSPORT; MEMBRANE; LIPOPROTEIN; SIGNAL.
FT SIGNAL 1 22 PROBABLE.
FT CHAIN 23 660 OLIGOPEPTIDE-BINDING PROTEIN ALIA.
FT LIPID 23 23 N-ACYL DIGLYCERIDE (PROBABLE).
FT VARIANT 7 7 L -> F (IN STRAINS NCTC 11906, SP-VA92,
FT VARIANT 27 27 SP-496, AND SP-VA96).
FT VARIANT 30 30 G -> D (IN STRAINS NCTC 11906, SP-VA92,
FT VARIANT 30 30 SP-496 AND SP-VA96).
FT VARIANT 166 166 A -> T (IN STRAINS NCTC 11906, SP-VA92,
FT VARIANT 166 166 SP-496 AND SP-VA96).
FT VARIANT 246 246 E -> D (IN STRAINS NCTC 11906, SP-VA92,
FT VARIANT 368 368 I -> V (IN STRAINS NCTC 11906, SP-VA92,
FT VARIANT 559 559 SP-496 AND SP-VA96).
FT VARIANT 612 612 L -> I (IN STRAINS NCTC 11906, SP-VA92,
FT VARIANT 612 612 SP-496 AND SP-VA96).
FT CONFLICT 18 19 T -> A (IN STRAINS NCTC 11906, SP-VA92,
FT CONFLICT 137 137 TT -> A (IN STRAIN SP-496).
FT CONFLICT 420 420 L -> P (IN REF. 3).
FT CONFLICT 420 420 A -> R (IN REF. 1).
SO SEQUENCE 660 AA; 73002 MW; 7FCAB451 CRC32;

Query Match 5.5%; Score 118.5; DB 1; Length 660;
Best Local Similarity 21.08; Pred. No. 2;
Matches 93; Conservative 65; Mismatches 170; Indels 115; Gaps 21;

QY 12 SSATYATSLASIIAFVA-AGCGQTSGSTSDSKPOAETLKHVNSDSIRIALTPDNPW 70
DB 3 SSKLLALAGVTLAATTAAAC-----SGSGSSAKGE-KTFSY-----IYETDPDN--- 47
QY 71 ISAKDIISYVDETEATSTITKN-----QDAQNWLTOQANLSPAPKGFIIAPENG 123
DB 47 -----LNYLTAKAATANITSNVVDGLLENDRYGNFVPSMA-----EDWS 86
QY 124 GVGTAVENTIADGPIVAYDRITGSDKYDWSYDFNEKVGELQGLSLAAGLLCKEDGAF 183
DB 87 -----VSKDGL---TYTITRKDAK--WYTS-EGEYAAVKAQDFVTGLKYAADKKS 132
QY 184 DSIDOMNEYLK---SHMPOETISFYTTA-GSQDDNNNSQYFYNGAMKVLKELMKNQKII 239
DB 133 DALYLQVESIKGLDAYVKGKIDFESQVGIKALDEQTVQYTLN-----KPSFWSKTTMG 187
QY 240 DLSPEGENAVYVPGWNYGTAGORTQSFLTINKDPA-----CGNKIKAVGSKPASIFKGL 294
DB 188 VLAPVNEEFNLKGGDDFAK-----TDPSSLLYNGPYLLKSIIVTKSSVEF--- 233
QY 295 APNDGMAEO---AATKLEGFDTQKIFVTRDYNDKAKT-----FIKGDQ 338
DB 233 AKNPYWDKNVHDVKVLSFDWQDQTSKPAENFDGSLTAARLYPTSPASFAELEKSMKD 292
QY 339 NMTYKPKVLGVAVELVRLVIAKKNKASREVENELAKLPN-----ISFKYDQNTYK 393
DB 293 NIVTQDSITVLVGTNIDROSRYKTSQTSDEQKASTKALLNDRQAIATAGFDRYAYA 352
QY 394 VQ-----GNINILVSPVIV 409
DB 353 SOLNGQTGASKILRNLFVPTTFV 375

RESULT 9
YTFQ_ECOLI
ID YTFQ_ECOLI
AC P39325;

DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ABC TRANSPORTER PERIPLASMIC BINDING PROTEIN YTFQ PRECURSOR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE; 95334362.
RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
RA BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL NUCLEIC ACIDS RES. 23:2105-2119(1995).
RN [2]
RP SEQUENCE OF 22-33.
RC STRAIN-K12 / EMG2;
RX MEDLINE; 97443975.
RA LINK A.J., ROBISON K., CHURCH G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL ELECTROPHORESIS 18:1259-1313(1997).
CC -|- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -|- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC RECEPTOR FAMILY 2.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14003; G537069; -
CC EMBL: AE000494; G1790674; -
CC ECGENE; EG12517; YTFQ_BP_like; 1.
CC PFAM; PF00532; Peripla_BP_like; 1.
CC HSSP; P02925; IURP.
CC TRANSPORT; PERIPLASMIC; SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 318 ABC TRANSPORTER PERIPLASMIC BINDING
FT PROTEIN YTFQ.
FT SEQUENCE 318 AA; 34344 MW; A2BE3935 CRC32;

Query Match 5.5%; Score 118; DB 1; Length 318;
Best Local Similarity 20.4%; Pred. No. 0.82;
Matches 78; Conservative 48; Mismatches 133; Indels 124; Gaps 13;

QY 7 KFLYSATYATSLASIIAFVAAGCGQTSGSTSDSKPOAETLKHVNSDSIRIALTPD 66
DB 3 KLLIIVSAVA-AMSSMALAAPLTGVFSQVSGSG----- 37
QY 67 NPMWISAQKDIISYVDETEATSTITKNQDAQNWL-TOQANLSPAPKGFIIAPENG 125
DB 37 ---WRAETNVKASEAKRGITLKIADGQKQENQKAVRSFVAQGVDAIFAPVATGW 93
QY 126 GTAVNTIADGPIVAYDRITGSDKY-----DWYVSFDNEK---VG 164
DB 94 EPIVLEAKDAEIPVFLDRSDIVKDKSLYMTVTADNILEGLIGDMLKVEVNGKFCNV 153
QY 165 ELQGLSLAAGLLCKEDGAFDSI-----DOMNEYKLSHMPQETISFYTTAGSQDDN 214
DB 154 ELQGTGASVAIDRKKGFAGAEIAKNAPNKIIRSQSGDFTRSKGVNESFI-----KAEN 208
QY 215 NSQ-----FFYN-----GAMKVLKELMKNQKIIIDLSPEGENAVYVPGWNYGTAGORIQ 264
DB 209 NGKNICMVYAHNDMDVIGAIQAIKAGLPGKDIILTGSIDGVVDIY----- 255
```

QY 265 SFLTKNDPAGGNKIKAVGSKPASIFKGLAPN-DGMAEQAIKTKLEGFTDKIFVTRQ 323
Db 255 -----KAMDGGEANASV---ELTPNAGPAFDALEKIKYKDGTMPEKLTITS 298
QY 324 DYNDKAKTFIKDGDQNTWTKPD 346
Db 299 -----TLTLPD 304
RESULT 10
RBSB_ECOLI STANDARD; PRT; 296 AA.
AC P02925;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE D-RIBOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR.
GN RBSB OR RBSB OR PRLB.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 84032513.
RA GROARKE J.M., MAHONEY W.C., HOPE J.N., FURLONG C.E., ROBB F.T.,
ZALKIN H., HERMODSON M.A.;
RT "The amino acid sequence of D-ribose-binding protein from Escherichia coli K12.";
RT J. BIOL. CHEM. 258:12952-12956(1983).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12;
RA MAUZY C.A.;
RL SUBMITTED (FEB-1986) TO EMBL/GENBANK/DBJ DATA BANKS.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MGI655;
RX MEDLINE: 93315143.
RA BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication.";
RL GENOMICS 16:551-561(1993).
[4]
RN SEQUENCE OF 1-12 FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 86224050.
RA BELL A.W., BUCKEL S.D., GROARKE J.M., HOPE J.N., KINGSLEY D.H.,
HERMODSON M.A.;
RT "The nucleotide sequences of the rbsD, rbsA, and rbsC genes of Escherichia coli K12.";
RL J. BIOL. CHEM. 261:7652-7658(1986).
[5]
RN SEQUENCE OF 270-296 FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 86224052.
RA HOPE J.N., BELL A.W., HERMODSON M.A., GROARKE J.M.;
RT "Ribokinase from Escherichia coli K12. Nucleotide sequence and overexpression of the rbsK gene and purification of ribokinase.";
RL J. BIOL. CHEM. 261:7663-7668(1986).
[6]
RN SEQUENCE OF 26-37.
RC STRAIN-K12 / EMG2;
RX MEDLINE: 97443975.
RA LINK A.J., ROBISON K., CHURCH G.M.;
RT "Comparing the predicted and observed properties of proteins encoded in the genome of Escherichia coli K-12.";
RL ELECTROPHORESIS 18:1259-1313(1997).
[7]
RN X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE: 92260530.
RA MOWBRAY S.L., COLE L.B.;
RT "1.7-A X-ray structure of the periplasmic ribose receptor from Escherichia coli.";

RL J. MOL. BIOL. 225:155-175(1992).
[8]
RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE: 95074014.
RA BJORKMAN A.J., BINNIE R.A., ZHANG H., COLE L.B., HERMODSON M.A.,
MOWBRAY S.L.;
RT "Probing protein-protein interactions. The ribose-binding protein in bacterial transport and chemotaxis.";
RL J. BIOL. CHEM. 269:30206-30211(1994).
[9]
RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE: 98312419.
RA BJORKMAN A.J., MOWBRAY S.L.;
RT "Multiple open forms of ribose-binding protein trace the path of its conformational change.";
RL J. MOL. BIOL. 279:651-664(1998).
CC -!- FUNCTION: INVOLVED IN THE HIGH-AFFINITY D-RIBOSE MEMBRANE TRANSPORT SYSTEM AND ALSO SERVES AS THE PRIMARY CHEMORECEPTOR FOR CHEMOTAXIS.
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING RECEPTOR FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K00511; G147519;
DR EMBL: M13169; G147515;
DR EMBL: L10328; G290601;
DR EMBL: AE000452; G1790192;
DR FIR: A03425; JGECR.
DR PDB: 2DRI; 26-JAN-95.
DR PDB: 1DBP; 15-OCT-94.
DR PDB: 1DRJ; 26-JAN-95.
DR PDB: 1DRK; 26-JAN-95.
DR PDB: 1URP; 17-JUN-98.
DR PDB: 1BA2; 15-JUL-98.
DR ECO2DBASE: H027.9; 6TH EDITION.
DR ECOGENE: EG10815; RBSB.
DR PFAM: PF00532; Peripla_BP_like; 1.
KW TRANSPORT; SUGAR TRANSPORT; PERIPLASMIC; CHEMOTAXIS; SIGNAL;
KW 3D-STRUCTURE.
FT SIGNAL 1 25
FT CHAIN 26 296 D-RIBOSE-BINDING PERIPLASMIC PROTEIN.
FT STRAND 28 33
FT HELIX 39 55
FT TURN 56 56
FT TURN 58 63
FT TURN 65 66
FT TURN 68 78
FT TURN 79 82
FT TURN 83 88
FT TURN 93 96
FT HELIX 97 105
FT TURN 106 107
FT STRAND 110 113
FT STRAND 124 128
FT HELIX 130 145
FT TURN 147 148
FT STRAND 150 155
FT TURN 158 159
FT HELIX 161 177
FT TURN 178 178
FT STRAND 180 186
FT TURN 188 189
FT HELIX 191 204
FT TURN 206 207
FT STRAND 210 213

RA HOGG R.W., VOELKER C., VON CARLOWITZ I.;
RT "Nucleotide sequence and analysis of the *mgI* operon of *Escherichia*
RT coli K12.";
RL MOL. GEN. GENET. 229:453-459(1991).
RN [2]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF GLY-74.
RC STRAIN=K12;
RX MEDLINE; 87286407.
RA SCHOLLE A., VREEMANN J., BLANK V., NOLD A., BOOS W., MANSON M.D.;
RT "Sequence of the *mgIB* gene from *Escherichia coli* K12: comparison of
RT wild-type and mutant galactose chemoreceptors.";
RL MOL. GEN. GENET. 208:247-253(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., ROBE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK R.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL SCIENCE 277:1453-1474(1997).
RN [5]
RP SEQUENCE OF 24-332.
RX MEDLINE; 81168234.
RA MAHONEY W.C., HOGG R.W., HERMODSON M.A.;
RT "The amino acid sequence of the D-galactose-binding protein from
RT *Escherichia coli* B/r.";
RL J. BIOL. CHEM. 256:4350-4356(1981).
RN [6]
RP REVISIONS TO 179-180.
RA MAHONEY W.C., HOGG R.W., HERMODSON M.A.;
RL SUBMITTED (NOV-1982) TO THE PIR DATA BANK.
RN [7]
RP SEQUENCE OF 1-58 FROM N.A.
RX MEDLINE; 83291030.
RA SCRIPTURE J.B., HOGG R.W.;
RT "The nucleotide sequences defining the signal peptides of the
RT galactose-binding protein and the arabinose-binding protein.";
RL J. BIOL. CHEM. 258:10853-10855(1983).
RN [8]
RP SEQUENCE OF 24-47.
RX STRAIN=K12 / EMG2;
RX MEDLINE; 97443975.
RA LINK A.J., ROBISON K., CHURCH G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of *Escherichia coli* K-12.";
RL ELECTROPHORESIS 18:1259-1313(1997).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE; 83169767.
RA VYAS N.K., VYAS M.N., QUIJOCHO F.A.;
RT "The 3-A resolution structure of a D-galactose-binding protein for
RT transport and chemotaxis in *Escherichia coli*.";
RL PROC. NATL. ACAD. SCI. U.S.A. 80:1792-1796(1983).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE; 87258173.
RA VYAS N.K., VYAS M.N., QUIJOCHO F.A.;
RT "A novel calcium binding site in the galactose-binding protein of
RT bacterial transport and chemotaxis.";
RL NATURE 327:635-638(1987).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE; 89058638.
RA VYAS N.K., VYAS M.N., QUIJOCHO F.A.;
RT "Sugar and signal-transducer binding sites of the *Escherichia coli*

galactose chemoreceptor protein.";
RL SCIENCE 242:1290-1295(1988).
RN [12]
RP FUNCTION: THIS PROTEIN IS INVOLVED IN THE ACTIVE TRANSPORT OF
RT GALACTOSE AND GLUCOSE. IT PLAYS A ROLE IN THE CHEMOTAXIS TOWARDS
CC THE TWO SUGARS BY INTERACTING WITH THE TRG CHEMORECEPTOR.
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -!- THE CALCIUM-BINDING SITE IS STRUCTURALLY SIMILAR TO THAT OF
CC EF-HAND PROTEINS, BUT IS IN TWO PARTS, WITH THE LAST CALCIUM
CC LIGAND PROVIDED BY GLU-228.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC RECEPTOR FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M59444; G146853; -
DR EMBL; X05646; G762931; -
DR EMBL; K00419; G146857; -
DR EMBL; U00007; G405891; -
DR EMBL; AE000304; G1788473; -
DR PIR; A03427; JGEGC.
DR PIR; Q00294; LFECMB.
DR PIR; A37277; A37277.
DR PDB; 2G8P; 15-JAN-92.
DR PDB; 1GLG; 31-MAY-94.
DR SWISS-2DPAGE; P02927; COLI.
DR ECO2DBASE; C029.8; 6TH EDITION.
DR ECOGENE; EG10593; MGLB.
DR PFAM; PF00532; PeriPla_BP_like; 1.
KW TRANSPORT; SUGAR TRANSPORT; PERIPLASMIC; CALCIUM-BINDING; CHEMOTAXIS;
KW SIGNAL; 3D-STRUCTURE.
FT SIGNAL 1 23
FT CHAIN 24 332
FT CA_BIND 157 165
FT CA_BIND 227 228
FT MUTAGEN 97 97
FT SITE 97 97
FT STRAND 26 32
FT TURN 35 36
FT HELIX 38 51
FT TURN 52 53
FT TURN 55 56
FT STRAND 57 63
FT TURN 65 66
FT HELIX 68 80
FT TURN 81 82
FT STRAND 85 88
FT HELIX 93 104
FT TURN 105 107
FT STRAND 110 113
FT HELIX 119 123
FT TURN 124 124
FT TURN 126 127
FT STRAND 128 132
FT HELIX 135 152
FT HELIX 154 156
FT TURN 158 159
FT STRAND 164 170
FT TURN 173 174
FT HELIX 176 191
FT TURN 192 193
FT STRAND 196 203
FT TURN 205 206
FT HELIX 208 218
FT TURN 219 220
FT TURN 222 229
FT STRAND 229 232

```

FT HELIX      235      247
FT TURN      248      249
FT TURN      251      252
FT STRAND    255      256
FT STRAND    259      259
FT HELIX     262      270
FT TURN      271      271
FT STRAND    275      278
FT HELIX     281      296
FT TURN      297      298
FT TURN      301      304
FT STRAND    310      310
FT TURN      311      312
FT STRAND    313      315
FT STRAND    319      321
FT TURN      323      325
FT HELIX     326      328
FT TURN      329      329
SQ SEQUENCE  332 AA; 35712 MW; 7D352971 CRC32;

Query Match      5.4%; Score 116; DB 1; Length 332;
Best Local Similarity 21.2%; Pred. No. 1.2;
Matches 85; Conservative 67; Mismatches 157; Indels 92; Gaps 21;

QY 5 LRKKFLYSIAIATSLASIIAFVAGCGQTSGTSKPKQAETLKHKVSNDISRIALTD 64
   : ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 1 MNKKVLTLSAV---MASMLFGAAHAADTRIGVT-----IYKY----- 36

QY 65 PDNPRWISAQKDIISYDVEATEATSTTKNQDAQNMLTQAN-----LSPAPKGFIIAPE 120
   || : || : || : || : || : || : || : || : || : || : || : ||
Db 36 DDN--FMSVVRKALE--QDAKAAPDVQLLMSNDQSQKNDQIDVLLAKGVKALAINLV 91

QY 121 NGSGVGTAVNTIADKGIPIVAYDLIT--GSDKYD--WYVSFDEKVGELQGLSLAAGLL 176
   : : : || : : : || : : : || : : : || : : : || : : : || : :
Db 92 DPAAAGTIVIEKARQNVVVFVFNKPKSKALDSYDKAYVGTGTSKESGIIQG-----DLI 146

QY 177 GKEDGAFSDIDOMNEYLKSHMPQETISFYTAG--SQDDNNSQYFYNGAMKVLKEL---- 231
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 147 AKHWAANQGWLD-LNK-----DQIQIVLLKGEPGHPDAEARTY-----VIKELNDKG 193

QY 231 MKNQNKIIDLSPGENAVVPGHNYCTAGRIOSFLTINKDPAGNKKIYKAVGSKPASIF 290
   : | : || : || : || : || : || : || : || : || : || : || : ||
Db 194 IKTEQLQL-----DTAM-----WDTAQAKDKMDAWLS-----GNANKIEW----- 231

QY 291 KGFLAPNDGMAEQAITKLKLEGFTQKIFVTRODYNDKAKTFIKDQDNMTIYKPKDKVLG 350
   : || || || : || : || : || : || : || : || : || : || : || :
Db 231 ---IANDAMAMGAVEALKAHNKSSIPVGV--DALPEALALYKSGALACTVLNDANNOA 285

QY 351 KVAVEVLRLVLIANKNKASRS--EVENELKALPNISKFYDN 389
   | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 286 KATFDLAKNLADGKGAADGTNWKIDNKV--VRVPYVGVYDKDN 325

RESULT 13
DGAL_SALTY
ID DGAL_SALTY STANDARD; PRT; 332 AA.
AC P23905;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE D-GALACTOSE-BINDING PERIPLASMIC PROTEIN PRECURSOR (GBP) (D-GALACTOSE/
GN MGLB.
OS SALMONELLA TYPHIMURIUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE: 89112167.
RA BENNER-LUGER D., BOOS W.;
RT "The mglB gene of Salmonella typhimurium LT2; promoter analysis

```

```

RT by gene fusions and evidence for a divergently oriented gene coding
RT for the mgl repressor."
RL MOL. GEN. GENET. 214:579-587(1988).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE: 83238396.
RA MOWBRAY S.L., PETSKO G.A.;
RT "The x-ray structure of the periplasmic galactose binding protein
RT from Salmonella typhimurium at 3.0-A resolution.";
RL J. BIOL. CHEM. 258:7991-7997(1983).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE: 94016588.
RA ZHOU J.Y., FLOCCO M.M., MOWBRAY S.L.;
RT "The 1.7 A refined x-ray structure of the periplasmic
RT glucose/galactose receptor from Salmonella typhimurium.";
RL J. MOL. BIOL. 233:739-752(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE: 94179303.
RA FLOCCO M.M., MOWBRAY S.L.;
RT "The 1.9 A x-ray structure of a closed unliganded form of the
RT periplasmic glucose/galactose receptor from Salmonella typhimurium.";
RL J. BIOL. CHEM. 269:8931-8936(1994).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE ACTIVE TRANSPORT OF
CC GALACTOSE AND GLUCOSE. IT PLAYS A ROLE IN THE CHEMOTAXIS TOWARDS
CC THE TWO SUGARS BY INTERACTING WITH THE TRG CHEMORECEPTOR.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.
CC -1- THE CALCIUM-BINDING SITE IS STRUCTURALLY SIMILAR TO THAT OF
CC EF-HAND PROTEINS, BUT IS IN TWO PARTS, WITH THE LAST CALCIUM
CC LIGAND PROVIDED BY GLU-228.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC RECEPTOR FAMILY 2.
DR PIR; S29390; S29390.
DR PDB; 3GBP; 15-JAN-93.
DR PDB; 1GCA; 31-JAN-94.
DR PDB; 1GCG; 31-MAY-94.
DR PFAM; PF00532; PeriPla_BP like; 1.
KW TRANSPORT; SUGAR TRANSPORT; PERIPLASMIC; CALCIUM-BINDING; CHEMOTAXIS;
KW SIGNAL; 3D-STRUCTURE.
FT SIGNAL 1 23
FT CHAIN 24 332
FT CA_BIND 157 165
FT CA_BIND 227 228
FT SITE 97 97
FT STRAND 26 32
FT TURN 35 36
FT HELIX 38 52
FT TURN 53 53
FT TURN 55 56
FT STRAND 57 63
FT TURN 65 66
FT HELIX 68 80
FT TURN 81 82
FT STRAND 85 88
FT HELIX 93 105
FT TURN 106 107
FT STRAND 110 113
FT HELIX 119 123
FT TURN 124 124
FT TURN 126 127
FT STRAND 128 132
FT HELIX 135 152
FT TURN 154 156
FT TURN 158 159
FT STRAND 164 170
FT TURN 173 174
FT HELIX 176 191
FT TURN 192 193
FT STRAND 196 203
FT TURN 205 206

```

```
FT HELIX 208 219
FT TURN 220 220
FT TURN 222 223
FT TURN 224 226
FT STRAND 229 232
FT HELIX 235 247
FT TURN 248 249
FT TURN 251 252
FT STRAND 255 256
FT STRAND 259 259
FT HELIX 262 270
FT TURN 271 271
FT STRAND 275 278
FT HELIX 281 296
FT TURN 297 298
FT TURN 301 304
FT STRAND 310 310
FT TURN 311 312
FT STRAND 313 315
FT STRAND 319 321
FT TURN 323 325
FT HELIX 326 328
FT TURN 329 329
SQ SEQUENCE 332 AA; 35747 MW; 3C67EE1D CRC32;

Query Match 5.3%; Score 113; DB 1; Length 332;
Best Local Similarity 20.3%; Pred. No. 1.8;
Matches 84; Conservative 61; Mismatches 152; Indels 116; Gaps 19;

QY 5 LRKFLYSIAIYATSLASIAIFVAAGCGQTESGST-----SUSKQP 45
D 1 MNKKVLTSAV-----MASLLFGAHAHAADTRIGVTIYKYDDNFMSVVRKAIEKDGKSAPD 56
QY 46 AETLKHVSNDSIRIALTDPNPRWISAOKDILSYVDETEAATITKNDQANNWLTTQ 105
D 57 VOLLNDSQNDQ-----SKQNDQIDVL-----79
QY 106 ANLSPAPKGFIIAPENGSGVGTAVNTIADRGIPVAYDRLIT--GSKDYD--WVVSFDNE 161
D 79 --LAKGVKALAINLVDPAAGTVIEKARGQNPVVFENKEPSRKALDSYDKAYVVGTDK 136
QY 162 KYGELQGLSLAAGLLGKEDGAFSDIDONNEYLKSHMPOETISFYTIAG--SQDDNNSQYF 219
D 137 ESGVIQG-----DLIAKHQWQAGWD--LNKDGK-----IQYVLLKGEPGAPDAEARTT 183
QY 220 YNGAMKVKELM--KNSONKIIDLSPENENAVYVPGWNYGTAGORIQSFLTINKDPAGGNK 278
D 184 Y-----VVKELNDKGIOTEQAL-----DTAM-----WDTAQAKKMDAWLS-----GPNANK 226
QY 279 IKAVGSRPASIFKGLAPNDGMAEQATIKLKGFTQKIFVTRQDYNDKAKTFIKDGDQ 338
D 227 IEVV-----IANDAMANGAVEALKAHNKSSIPVFGV--DALPEALALVRKSGAM 273
QY 339 NWTIYKPKVLGVAVELRVLAKNKASRS--EVENEKAKLPNISFYDN 389
D 274 AGTVLNDANNQAKATFDLAKNLAEGKAADGTWIKENKI--VRVPYVGVDKDN 325

RESULT 14
S230_PLAFO STANDARD; PRT; 3135 AA.
AC Q08372;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE TRANSMISSION-BLOCKING TARGET ANTIGEN S230 PRECURSOR.
GN S230.
OS PLASMIDIUM FALCIPARUM (ISOLATE NF54), AND
OS PLASMIDIUM FALCIPARUM (ISOLATE 3D7).
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMIDIUM.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
```

```
RC STRAIN-ISOLATE NF54;
RX MEDLINE; 93241227.
RA WILLIAMSON K.C., CRISCIO M.D., KASLOW D.C.;
RT "Cloning and expression of the gene for Plasmodium falciparum
transmission-blocking target antigen, Pfs230.";
RL MOL. BIOCHEM. PARASITOL. 58:355-358(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE 3D7;
RA BHATTI S., ALANO P., LUO C., HANSRA S., AIKAWA M., CARTER R.,
RA ELLIOTT J.F.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: SURFACE PROTEIN.
CC -!- DEVELOPMENTAL STAGE: GAMETOCYTE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L08135; G294176; -
DR EMBL; L04162; G605631; -
DR PIR; A48584; A48584.
KW MEMBRANE; REPEAT; MALARIA; ANTIGEN; SIGNAL.
FT SIGNAL 1 20
FT CHAIN 21 3135
FT
FT DOMAIN 280 304
FT DOMAIN 379 410
FT REPEAT 379 382
FT REPEAT 383 386
FT REPEAT 387 390
FT REPEAT 391 394
FT REPEAT 395 398
FT REPEAT 399 402
FT REPEAT 403 406
FT REPEAT 407 410
FT DOMAIN 411 442
FT REPEAT 411 418
FT REPEAT 419 426
FT REPEAT 427 434
FT REPEAT 435 442
SQ SEQUENCE 3135 AA; 363213 MW; FCAEF748 CRC32;

Query Match 5.2%; Score 112.5; DB 1; Length 3135;
Best Local Similarity 20.1%; Pred. No. 40;
Matches 87; Conservative 62; Mismatches 168; Indels 115; Gaps 18;

QY 35 ESGSTSDSPQAETLKHVSNDSIRIALTDPNPRWISAOKDIL--SYVDETEAAT--S 89
D 312 ESGDETEQLQEEH-QEEVGAESSEESDEDEDSVEARDGMIRVDEYEDQDGDYDS 370
QY 90 TITKNODANNLTTQANLSPAPKGFIIAPENGSGVGTAVNTIADRGIPVAYDRLITGS 149
D 371 TI-KNEDVDEE-----VGEVGEVGEVGEVGEVGEV-----GE 407
QY 150 DRYDWTYVFDNEKRYGELQGLSLAAGL---LGKEDGAFSDIDONNEYLKSHMPOETISFYT 206
D 408 EV-----GEEVGEVGEVGEVGEVGEVGEVGEVGEVGEVGEVGEVGEVGEVGEVGEV 458
QY 207 IAGSQDDNNSQYFYNGAMKVKELMKNKNSONKIIDLSPENENAVYVPG-----254
D 459 DEEEKDGESSEFYKRS-----EVDKTLDFKFTIE-GGEGDDVYKVDGSKVLLDDDTISR 512
QY 254 -----WNYGTAGORIQSFLTINKDPAGGNKIKAVGSRPASIFKGLAPNDGMAEQAI 305
D 513 SKKHTARDEIGEYGEVGEVGEV-----GENVIKII-----RSVLSQSGALPSVGVDEL 559
```


Search completed: November 13, 1999, 10:33:27
Job time: 5177 sec

[illegible]

GenCore version 4.5

Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:05:01 ; Search time 139.86 Seconds

(without alignments)
184.376 Million cell updates/sec

Title: US-08-913-430-2

Perfect score: 2143

Sequence: 1 MKMLRKFLYSAYATSL.....NTILSPVIVTKANVNDPDA 419

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPREMBL_10:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	337	15.7	366	2 068456	thermoanaer
2	222	10.4	371	2 050503	streptomyce
3	128	6.0	289	2 082843	tetragenoco
4	127.5	5.9	661	2 052230	streptococc
5	124	5.8	630	10 Q40161	Q40161 lycopersico
6	120	5.6	961	2 Q49551	Q49551 mycoplasma
7	119.5	5.6	661	2 033900	Q33900 shewanella
8	119.5	5.6	3134	5 Q25994	Q25994 plasmodium
9	118	5.5	947	2 086487	Q86487 staphylococ
10	117	5.5	1471	5 Q18245	Q18245 caenorhabdi
11	115.5	5.4	470	2 007041	Q07041 streptococc
12	113.5	5.3	2353	2 P71401	P71401 haemophilus
13	113.5	5.3	1119	2 051228	Q51228 borrelia bu
14	112.5	5.2	667	2 054232	Q54232 streptococc
15	112.5	5.2	1315	2 086488	Q86488 staphylococ
16	111	5.2	3016	2 P73590	P73590 synechocyst
17	111	5.2	4152	2 Q92HL3	Q92HL3 haemophilus
18	110.5	5.2	345	2 Q49738	Q49738 mycobacteri
19	108.5	5.1	451	2 P71070	P71070 bacillus su
20	108.5	5.1	1055	4 Q43477	Q43477 homo sapien
21	108	5.0	1302	2 Q49547	Q49547 mycoplasma
22	108	5.0	1335	2 Q33635	Q33635 staphylococ
23	108	5.0	1127	12 Q9VIT6	Q9VIT6 melanoplus
24	107.5	5.0	1231	10 Q82276	Q82276 arabidopsis
25	106.5	5.0	668	2 Q51252	Q51252 borrelia bu
26	106	4.9	1098	2 Q48152	Q48152 haemophilus
27	106	4.9	2269	5 Q26223	Q26223 plasmodium
28	105	4.9	1078	2 Q49529	Q49529 mycoplasma
29	104.5	4.9	718	1 Q58791	Q58791 pyrococcus

```

30 104.5 4.9 334 2 068119 068119 rhodobacter
31 104.5 4.9 1054 2 0861X4 0861X4 pseudocalter
32 104 4.9 1463 2 086919 086919 staphylococ
33 104 4.9 2401 5 026216 026216 plasmodium
34 104 4.9 577 5 035995 035995 vairimorpha
35 104 4.9 852 12 073350 073350 human immun
36 104 4.9 852 12 073352 073352 human immun
37 104 4.9 852 12 073354 073354 human immun
38 103.5 4.8 321 2 052943 052943 rhizobium m
39 103 4.8 1288 4 095752 095752 homo sapien
40 102.5 4.8 865 10 024320 024320 phaseolus v
41 102 4.8 4199 2 074440 074440 synechocyst
42 102 4.8 639 10 023133 023133 arabidopsis
43 101.5 4.7 614 2 046518 046518 bacteroides
44 101.5 4.7 1165 2 067411 067411 aquifex aeo
45 101.5 4.7 869 5 094174 094174 caenorhabdi

```

ALIGNMENTS

```

RESULT 1
068456 PRELIMINARY; PRT; 366 AA.
ID 068456;
AC 068456;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE D-XLOSE-BINDING PROTEIN.
GN XILF.
OS Thermoanaerobacter ethanolicus (Clostridium thermohydrosulfuricum).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Thermoanaerobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-39E;
RA ERBEZNIK M., STROBEL H.J., DAWSON K.A., JONES C.R.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043466; AAC38794.1;
DR PFAM; PF00532; Peripla_Bp_like; 1.
SQ SEQUENCE 366 AA; 39805 MW; D7AB906D CRC32;

```

Query Match 15.7%; Score 337; DB 2; Length 366;

Best Local Similarity 28.1%; Pred. No. 1.1e-13;

Matches 118; Conservative 62; Mismatches 156; Indels 84; Gaps 16;

QY 1 MKMLRKFLYSAYATSLAIIAFVAAAGCGQTSGTSDSKPQAEHLKHVNSDSIRI 60

Db 1 MFKVSKKF-----ALLIVAMFT-LGIILSGCATTONNQSSSTTPSSNESTKAPEKIKI 55

QY 61 ALTDPD---PWIISAQKD---LIIVDETEATSTITKNQDAQNWLTPQANLSPAPKG 114

Db 56 GFS-MDNLALRW---QKDRDLFVKAQELVAELVQSANSDFQLQYSQCQLLAGIQV 111

QY 115 FIAPENGSGVGTAVNTIADKGIPIVAYDRLITGSDKYDWYVDFDNKVKYGLQGLSLAAG 174

Db 112 LVILPDGSAIPIVEEAHKAQKVLAYDRLIMNSD-LDYVYVDFDNKVKYGLQA----- 165

QY 175 LLGKEDGAFDSIDQWNEYLKSHMPOETTSFYTIAGSQDDNNQSOYFYNGAMVKVLELMKNS 234

Db 165 -----EATIKLVPKG--RYFLLEGSTPDNNAKLFYEGQMKVLPKLVLDKG 206

QY 235 QNKIIDLSPGENAVYVPCWNYGTAGRIQSFLTINKDPAGNKKIKAVGSKPASIFKGL 294

Db 207 DKIV-----GEO--WAKNWDNEAYKIMONALVAN-----NNKIDAV-----V 243

QY 295 APNDGMAEQATKLEGGEDTQKIFVTQDYNDKARTFKIDGQNNNTIYKPKVLGVAV 354

Db 244 AANDSTALGAIRALDEOGL-AGKIPISGQADLANCOLIVEGQKQMTVYKPIVLEATKAA 302

QY 355 EYLRVLIAKNKASRSEVENELKAKPLNITSFYDQTYKVGKNNINTILVSPVIVTKANY 414

Db 355 EYLRVLIAKNKASRSEVENELKAKPLNITSFYDQTYKVGKNNINTILVSPVIVTKANY 414

[illegible]

[illegible]

Db 451 QGEIYVFGDEBEKDEGESFTYEKS-----EVDKTDLFKFI-E-GGEGDYYVVDGSKVLL 504
QY 254 -----WNYGTAGRIQSFLTINKDPAGNKKIKAVGSKPASIFKGLAPN 297
Db 505 DDDTISRVSKKHTARGEGYGEAVED-----GENVIKII-----RSVLQSGALPS 551
QY 298 DGMASQAITKLEGFDTQIKIYTRQDY-----NDKAKTFIKGDQONMTIYKPKVLGKV 352
Db 552 VGVDELKIDLSYETTESDGTAVSDSYDKYASNNTKNEYVCDFTDQKPTESSPKVRKC 611
QY 353 AVELRVLIKAKNKSRSVEVELKAKLPNISFKYDQNY-----K 393
Db 612 EYKVEPLIKV-----IICPLGSEVKL-----IDNIEYVPKSPYVVLTKREPKLKEK 661
QY 394 VQGKNINTILVSPVIIVTKAN 413
Db 662 LLSKLIYGLLISPTVNEKEN 681
RESULT 9
ID O86487 PRELIMINARY; PRT; 947 AA.
AC O86487;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE SDRC PROTEIN.
GN SDRC.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN-NEWMAN;
RA FOSTER T.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NEWMAN;
RA FOSTER T.J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NEWMAN;
RA JOSEFSSON E., MCCREA K., NI EIDHIN D., O'CONNELL D.;
RT "Three new members of the serine-aspartate multigene family of
RT Staphylococcus aureus."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ005645; CAA06650.1; -
DR PFAM; PF00746; Gram_pos_anchor; 1.
SQ SEQUENCE 947 AA; 102888 MW; 89AB0255 CRC32;

Query Match 5.5%; Score 118; DB 2; Length 947;
Best Local Similarity 21.9%; Pred. No. 15;
Matches 86; Conservative 58; Mismatches 146; Indels 102; Gaps 22;

QY 44 POAEITLK-HKVSNDRIALTPDPRW--ISAQKDIISYVDTEAATSTIFKNQDAQN- 100
Db 386 PNAKFKIYEVDNQFVDSFTPTQSKLDVTDQFDVI-YSNDKTATVDMKGTSSNK 444
QY 100 NMLTQOANLSPAPKFIAPENGSGVGTAVNTIAKGIPIVAYDRLTGSDKYDNYVSD 159
Db 445 QVLIQQ-----VAYPDNSDNGKIDITLD-----TDKTYSHSNYS 482
QY 160 NEKVGELOGLSLAAGLLGKEDGAFDSIDOMNEYLKSHMPQETISFYTTAGSQDDNNQYF 219
Db 483 N-----VNGSSFANG-----DOKYNLGDYVWEDT-----NKDGKQDANEK 519
QY 220 YNGAMKVLKELMKNQNKIIDLSPGEGNAVY-VPGWNVYGTAGRIQSFLTINKDPAGNKK 278
Db 519 ---GIRGVTVILKDSNGKRLDRTTDENGKYQFTGLSNGTYSVFST-----PAGYTP 568
QY 279 IRA-VGSKPASIFKGLAPNDGMAEQAITK-----LKLEG--EDTQIKFVTRQDYNDKAKT 331
Db 569 TTANVGTDAA-----VDSDDLITGVVLIKADANMTLDSGFYKTPKSYGLDVIWYDSN-- 620
QY 332 FIKGDQONMTIYKPKVLGKAVAVELRVLIKAKNKSRSVEVELKAKLPNISFKYDQNT 391

Db 620 --KDGKQDST-----EKGIKGVKVLQN---EKGEVIGTETDENGKYRFDNL-----DSGK 666
QY 392 YKVOGKNINTILVSPVIIVTKANV-----DNPDA 419
Db 667 YKV-----IFEXPAGLTQGTNTTEDDKDA 691
RESULT 10
ID O18245 PRELIMINARY; PRT; 1471 AA.
AC O18245;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE C27B7.7 PROTEIN.
GN C27B7.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RA SEQUENCE FROM N.A.
RP MATTHEWS P.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SVALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WAYSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans".
RL Nature 368:32-38(1994).
DR EMBL; 254236; CAA90982.1; -
DR PFAM; PF00041; fn3; 7.
SQ SEQUENCE 1471 AA; 165507 MW; 1D71ECFC CRC32;

Query Match 5.5%; Score 117; DB 5; Length 1471;
Best Local Similarity 21.08; Pred. No. 32;
Matches 101; Conservative 75; Mismatches 208; Indels 96; Gaps 22;

QY 2 KMLR-----KKFLYSSAIYATSLA-SIIAFVAAGCGQTE-----SGSTSDSKP 44
Db 400 EKMVRIIMDTESFYGRVQAAATELPGIISDIVAMERTQIPISVESDLFGVYSATMVVNP 459
QY 45 QAETLKHVS-----NDSIRIALTPDPRWI-----SAQKDIISYVDETEAATS 89
Db 460 R-ETSIQCTARGPRPSISVAISDRKNASQVEVDWVSRLOATSSAGIVSAVHNFSLVTS 518
QY 90 TI-----TKNQDAQNMLTQOANL---SPAPKFIAPENGSGVGTAVNTIAKGIPIVAY 142
Db 519 KFHCRANK-SAGSNYSTMLKVKDPKPGDAPTOIQVLSNALDALVWVHSPFPNPSITY 577
QY 143 DRLITGSDKYD---WYVSFDNEKVELQGLSLAAGLLGKEDGAFDSIDOMNEYLKSHMPQ 199
Db 578 IVLVSNDDKEDKSTWLQYESNAKETQINRMLLPTGNLEKSTEFYFCVRKNA-AAGIGTIS 636
QY 200 ETISFYTTAGSQDD--NNSQIFYNGAMKVLKELMKNQNKIIDLSPGEGNAVYVPGWNYG 257
Db 637 SLISFITLNGPDSPPDNLKVLNEANQVI--VYWNTPNSTEVT--GYLIYITRDLISLS 692
QY 258 TAGQRIQSFLTINKDPAGNKKIKAVGSKPASIFK-----GFLAPNDGMAEQAI----- 306
Db 693 NDDYKNWQFVEMNNNTRYKFDLSVLGLKPKTFYRVRISGKNSHADGCPASVVEFFETAYSE 752

[illegible]

GenCore version 4.5
Copyright (C) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:59:43 ; Search time 104.22 Seconds
(without alignments)
5.682 Million cell updates/sec

Title: US-08-913-430-3
Perfect score: 122
Sequence: 1 AGXGQTESGSDSRPQATLKHV 25

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	120	98.4	25	1 W01033	Mycoplasma 46-48 k
2	119	97.5	419	1 R21829	Sequence of surfac
3	119	97.5	419	1 W01037	Mycoplasma 46-48 k
4	60	49.2	261	1 Y00032	Enterococcus faeca
5	52	42.6	240	1 Y00033	Enterococcus faeca
6	46	37.7	414	1 Y00529	Saccharomyces cere
7	45	36.9	1608	1 R98619	Borna disease viru
8	45	36.9	1711	1 R98605	Borna disease viru
9	45	36.9	621	1 W76213	Human ELL protein.
10	45	36.9	722	1 W93493	N. meningitidis st
11	45	36.9	707	1 W93496	N. meningitidis st
12	44.5	36.5	414	1 W94665	Streptococcus equi
13	44	36.1	543	1 R24022	Human promyelo-leu
14	44	36.1	741	1 P80136	Neisseria IgA-Prot
15	44	36.1	1164	1 W22469	Streptococcal C4a
16	44	36.1	543	1 W76985	Human EGR-1 protel
17	44	36.1	433	1 Y00166	Enterococcus faeca
18	44	36.1	409	1 Y00167	Enterococcus faeca
19	43	35.2	168	1 R13368	Heparin-binding gr
20	43	35.2	482	1 R89919	RPDL, a human foet
21	43	35.2	482	1 W29324	A novel histone de
22	43	35.2	342	1 Y00100	Enterococcus faeca
23	43	35.2	322	1 Y00101	Enterococcus faeca
24	42	34.4	273	1 W00161	GrpE encoded by AR
25	42	34.4	1429	1 W93941	Human brx protein.
26	41.5	34.0	531	1 R11829	FB-FB-SK fusion co
27	41.5	34.0	371	1 R10195	Streptokinase (1-3
28	41.5	34.0	414	1 R10194	Streptokinase enco
29	41.5	34.0	372	1 R10197	Streptokinase (1-3
30	41.5	34.0	374	1 R10198	Streptokinase (1-3
31	41.5	34.0	372	1 R10200	Streptokinase (1-3
32	41.5	34.0	414	1 W63120	Plasminogen-bindin
33	41.5	34.0	800	1 W21723	Modified streptoki
34	41.5	34.0	1194	1 W21724	Streptokinase/malt
35	41.5	34.0	1194	1 W21726	Streptokinase/malt
36	41.5	34.0	1181	1 W21727	Wild type plasmino
37	41.5	34.0	813	1 W21728	Modified streptoki
38	41.5	34.0	1194	1 W21725	Streptococcus equi
39	41.5	34.0	414	1 W94664	Native streptokina
40	41.5	34.0	414	1 Y01556	Human alpha 5 (IV)
41	41	33.6	772	1 R23873	S. cerevisiae Plc1
42	41	33.6	869	1 R53732	Cold acclimatizati
43	41	33.6	802	1 R56550	

ALIGNMENTS

```
RESULT 1
W01033
ID W01033 standard; Peptide; 25 AA.
AC W01033:
DT 19-JAN-1997 (first entry)
DE Mycoplasma 46-48 kDa protective antigen N-terminal peptide.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumonia;
   diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
FH Key Location/Qualifiers
FT misc_difference 3
FT /note= "undetermined amino acid"
PN W09628472-A1.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (UYME ) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
DR WPI; 96-433763/43.
PT Putative protective antigens against Mycoplasma - used for the
   detection, prevention or treatment of Mycoplasma infections, esp. M.
   hyopneumoniae in swine
PT Claim 12; Page 28; 43pp; English.
PS A 46-48 kDa putative protective antigen against Mycoplasma contains
   the N-terminal sequence given in W01033 and the internal CNBr
   fragments given in W01034-36. The antigen was isolated from
   Mycoplasma hyopneumoniae cells using antibody probes enriched with
   Mycoplasma-specific antibodies. Other protective antigens were
   also identified (see also W01024-32). A gene (r38241) coding for
   the 48 kDa antigen (W01037) was isolated from a genomic library.
   CC Protective antigens and antibodies can be used in vaccines for
   preventing or treating mycoplasma infections, partic. M.
   CC hyopneumoniae infections in swine. They can also be used for
   CC diagnosis. 25 AA;
SQ
```

Query Match 98.4%; Score 120; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.8e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AGXGQTESGSDSRPQATLKHV 25
   |||||
Db 1 AGXGQTESGSDSRPQATLKHV 25
```

```
RESULT 2
R21829
ID R21829 standard; Protein; 419 AA.
AC R21829;
DT 02-NOV-1992 (first entry)
DE Sequence of surface antigen 46kd.
KW Swine pneumonia; epidemic; diagnosis; therapy.
OS Mycoplasma hyopneumoniae.
EP 475185-A.
PD 18-MAR-1992.
PF 27-AUG-1991; 114335.
PR 27-AUG-1990; JP-224945.
PA (NIFL-) NIPPON FLOUR MILLS.
PI Seto Y, Futo S, Mitsuse S, Matsuo K, Tsuna M;
DR WPI; 92-089874/12.
DR N-PSDB; Q22042.
PT DNA and peptide of mycoplasma hyopneumoniae - useful for
   diagnosis and treatment of swine mycoplasma pneumonia
   Disclosure; Page 4-6 and pages 34-36; 45pp; English.
PS The inventors claim DNA encoding a surface antigen and primers used
   CC
```


CC The present sequence represents an antigenic polypeptide fragment
 CC isolated from *Enterococcus faecalis*. The present invention describes
 CC genes, proteins and antigenic polypeptides isolated from *E. faecalis*.
 CC The proteins can be used in vaccines for preventing or attenuating an
 CC infection caused by a member of the *Enterococcus* genus in an animal.
 CC They can also be used for detecting *Enterococcus* antibodies in a sample.
 CC The nucleotide sequences can be used for detecting *Enterococcus* nucleic
 CC acids. Products from the present invention can also be used for
 CC screening compounds to identify agonists and antagonists of *E. faecalis*
 CC protein activity.
 CC Sequence 240 AA;
 SQ

Query Match 42.6%; Score 52; DB 1; Length 240;

Best Local Similarity 52.6%; Pred. No. 1.3;

Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 GQTESGTSKPKQAEHLK 22

| : : | : | : | : | : | : |

Db 2 GNKKTASTNDSKPKQETKK 20

RESULT 6

W10529

ID W10529 standard; Protein; 414 AA.

AC W10529;

DT 02-APR-1997 (first entry)

DE Saccharomyces cerevisiae nucleolin like protein, NOL1.

KW Temperature: nucleolin; like; protein; NOL1; low; tolerance;

KW yeast; cold; induced; fermentation; expression; shock.

OS Saccharomyces cerevisiae (S288C).

PN US5470971-A.

PD 28-NOV-1995.

PF 11-MAR-1991; 667276.

PR 11-MAR-1991; US-667276.

PA (UYNE-) UNIV NEW JERSEY.

PI Inouye M, Kondo K;

DR N-PSDB; T60713.

DT Isolated stress-inducible yeast gene which encodes useful proteins

PT - which contribute to confer thermo or low temp. tolerance to

PT organisms

PS Disclosure: Columns 37-42; 60pp; English.

CC The present sequence is the Saccharomyces cerevisiae nucleolin

CC like protein, NOL1, which confers low temp. tolerance to organisms

CC in which it is expressed. Yeast transformed with the NOL1 gene, a

CC cold induced gene, can be fermented at lower than normal growth

CC temps. This enables valuable proteins, e.g. physiological active

CC proteins, to be expressed at lower than normal temps., avoiding

CC high temp. expression, which can be detrimental to full

CC physiological activity.

CC Sequence 414 AA;

SQ

Query Match 37.7%; Score 46; DB 1; Length 414;

Best Local Similarity 45.0%; Pred. No. 22;

Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 5 QTESGTSKPKQAEHLK 24

: : | : | : | : | : | : |

Db 65 ESSSSSSDSEAEATKKEE 84

RESULT 7

R98619

ID R98619 standard; Protein; 1608 AA.

AC R98619;

DT 10-DEC-1996 (first entry)

DE Borna disease virus polymerase.

KW Borna disease virus; BDV; G-protein; p57; nervous system disease;

KW neuro-psychiatric disease; schizophrenia; diagnosis; therapy;

KW vaccine; antibody.

OS Borna disease virus strain V.

Query Match 36.9%; Score 45; DB 1; Length 1711;

PN W09621020-A2.

PD 11-JUL-1996.

PF 05-JAN-1996; U00418.

PR 06-JAN-1995; US-369822.

PR 04-MAY-1995; US-434831.

PR 04-JAN-1996; US-582776.

PA (REGC) UNIV CALIFORNIA.

PI Briese T, Kliche S, Lipkin WI, Schneemann A, Schneider PA;

PI Stitz L;

DR WPI; 96-333995/33.

DR N-PSDB; T38104.

PT Borna disease virus (BDV) nucleotide and protein sequences - useful

PT for the diagnosis and treatment of infection and non-BDV related

PT neuro-logic and neuro-psychiatric disease

PS Claim 2; Fig 2; 186pp; English.

CC Borna disease virus (BDV) polymerase (R98619), or pol or p180, was

CC identified from an ORF on the virus antigenome strand (T38104).

CC The amino acid sequence for pol after splice modification is given

CC in R98605. Recombinant pol (recpol) can be expressed in transformed

CC host (partic. mammalian) cells. It is useful in assays for

CC detecting BDV infection and for diagnosing non-BDV related neurologic

CC and neuropsychiatric diseases. It may also be incorporated into

CC vaccines and used to raise anti-BDV antibodies.

CC Sequence 1608 AA;

SQ

Query Match

Best Local Similarity 36.9%; Score 45; DB 1; Length 1608;

Matches 10; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 7 ESGTSDSKP--QAEHLKHV 25

| | | | | : | : | : |

Db 1014 EKGYSIDARPSIQGGTLTHRL 1034

RESULT 8

R98605

ID R98605 standard; Protein; 1711 AA.

AC R98605;

DT 10-DEC-1996 (first entry)

DE Borna disease virus polymerase.

KW Borna disease virus; BDV; G-protein; p57; nervous system disease;

KW neuro-psychiatric disease; schizophrenia; diagnosis; therapy;

KW vaccine; antibody.

OS Borna disease virus strain V.

PN W09621020-A2.

PD 11-JUL-1996.

PF 05-JAN-1996; U00418.

PR 06-JAN-1995; US-369822.

PR 04-MAY-1995; US-434831.

PR 04-JAN-1996; US-582776.

PA (REGC) UNIV CALIFORNIA.

PI Briese T, Kliche S, Lipkin WI, Schneemann A, Schneider PA;

PI Stitz L;

DR WPI; 96-333995/33.

DR N-PSDB; T38103.

PT Borna disease virus (BDV) nucleotide and protein sequences - useful

PT for the diagnosis and treatment of infection and non-BDV related

PT neuro-logic and neuro-psychiatric disease

PS Claim 2; Page 131-135; 186pp; English.

CC Borna disease virus (BDV) polymerase (R98605), or pol or p180, was

CC identified from an ORF (T38103) on the virus antigenome strand

CC (T38104). The amino acid sequence of pol prior to splice

CC modification is given in R98619. Recombinant pol (recpol) can be

CC expressed in transformed host (partic. mammalian) cells. It is

CC useful in assays for detecting BDV infection and for diagnosing

CC non-BDV related neurologic and neuropsychiatric diseases. It may

CC also be incorporated into vaccines and used to raise anti-BDV

CC antibodies.

CC Sequence 1711 AA;

SQ

```

PT prevention and treatment of neisserial disease, e.g. meningitis
PS Claim 13; Page 87-89; 116pp; English.
CC This invention describes novel lactoferrin-binding protein B (LbpB)
CC strains of Neisseria meningitidis. The products of this invention can
CC be used for vaccinating humans against neisserial disease e.g.
CC meningitis. Antibodies raised against the proteins of the invention
CC can be used for diagnosing or treating neisserial disease in humans.
CC The lbpB polypeptides can also be used for identifying compounds which
CC inhibit the polypeptides.
SQ Sequence 722 AA;

Query Match          36.9%; Score 45; DB 1; Length 722;
Best Local Similarity 39.1%; Pred. No. 64;
Matches      9; Conservative    5; Mismatches   9; Indels    0; Gaps    0;

QY      1 AGXGOTESGSTSDSKPQAETLKH 23
        ||:|||:|||:|||:|||
DB      353 AGKQTETANASDPNLPALPSGKH 375

RESULT 11
W93496
ID W93496 standard; Protein; 707 AA.
AC W93496;
DE 11-JUN-1999 (first entry)
DE N. meningitidis strain 881607 LbpB protein.
DE LbpB; lactoferrin binding protein; vaccine; neisserial disease;
KW KWK; lactoferrin binding protein; diagnosis; treatment.
OS Neisseria meningitidis.
PN WO9009176-A1.
PD 25-FEB-1999.
PP 10-AUG-1998; E05117.
PR 05-FEB-1998; GB-002544.
PR 15-AUG-1997; GB-017423.
PA (UYUR-) RIJKSUNIV UTRECHT.
PA (TECH-) TECHNOLOGY FOUND TECHNOLOGIESTICHTING ST.
PI Pettersson-Fernholm AM, Tommassen JPM;
DR WPI; 99-190165/16.
DR N-PSDB; X23323.
PT New lactoferrin-binding protein B polynucleotides - obtained from

```

```

PS prevention and treatment of Neisserial disease, e.g. meningitis
PT Claim 13; Page 110-112; 116pp; English.
CC This invention describes novel lactoferrin-binding protein B (LbpB)
CC strains of Neisseria meningitidis. The products of this invention can
CC be used for vaccinating humans against Neisserial disease e.g.
CC meningitis. Antibodies raised against the proteins of the invention
CC can be used for diagnosing or treating Neisserial disease in humans.
CC The LbpB polypeptides can also be used for identifying compounds which
CC inhibit the polypeptides.
SQ Sequence 707 AA;

Query Match 36.9%; Score 45; DB 1; Length 707;
Best Local Similarity 39.1%; Pred. No. 63;
Matches 9; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 AGXGQTSGSTSDSKPOAETLKH 23
   || :||: : ||: : ||
Db 354 AGKQKTTANASDTNPALPSGKH 376

RESULT 12
W94665
ID ID W94665 standard; Protein; 414 AA.
AC W94665;
DT 04-MAY-1999 (first entry)
DE Streptococcus equisimilis mutant streptokinase K59E.
KW Streptococcus equisimilis H46A; streptokinase; mutant; fibrinolytic;
KW plasma clot; hydrolysis; haemolytic Streptococcus; plasminogen;
KW plasmin; serine protease; fibrin; blood clot; thrombolytic;
KW vascular thromboembolytic symptom; acute myocardial infarction;

```

Query Match	36.1%	Score 44;	DB 1;	Length 543;
Best Local Similarity	47.4%	Pred. No. 65;		

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	46	37.7	414	1	US-07-667-276A-4	Sequence 4, Appl
2	44	36.1	543	2	US-08-224-482-A	Sequence 4, Appl
3	44	36.1	1164	2	US-08-589-756-1	Sequence 1, Appl
4	43	35.2	482	1	US-08-528-255A-1	Sequence 1, Appl
5	43	35.2	482	2	US-08-717-365-1	Sequence 1, Appl
6	42	34.4	559	1	US-08-313-553-15	Sequence 15, Appl
7	41.5	34.0	1194	2	US-08-488-940-1	Sequence 1, Appl
8	41.5	34.0	1181	2	US-08-488-940-2	Sequence 2, Appl
9	41.5	34.0	813	2	US-08-488-940-3	Sequence 3, Appl
10	41.5	34.0	800	2	US-08-488-940-4	Sequence 4, Appl
11	41.5	34.0	1194	2	US-08-488-940-17	Sequence 17, Appl
12	41.5	34.0	1194	2	US-08-488-940-18	Sequence 18, Appl
13	41.5	34.0	413	2	US-08-759-599-12	Sequence 12, Appl
14	41	33.6	830	2	US-08-110-158-4	Sequence 4, Appl
15	41	33.6	802	2	US-08-007-107-4	Sequence 4, Appl
16	41	33.6	1167	2	US-08-589-756-2	Sequence 2, Appl
17	40.5	33.2	440	1	US-07-854-596B-15	Sequence 15, Appl
18	40.5	33.2	435	1	US-07-854-596B-19	Sequence 19, Appl
19	40.5	33.2	415	1	US-07-854-596B-26	Sequence 26, Appl
20	40.5	33.2	499	1	US-07-854-596B-28	Sequence 28, Appl
21	40.5	33.2	369	1	US-07-854-596B-31	Sequence 31, Appl
22	40.5	33.2	859	1	US-07-854-596B-35	Sequence 35, Appl
23	40.5	33.2	747	1	US-07-854-596B-40	Sequence 40, Appl
24	40.5	33.2	483	1	US-07-854-596B-43	Sequence 43, Appl
25	40.5	33.2	483	1	US-07-854-596B-47	Sequence 47, Appl
26	40	32.8	512	1	US-07-623-953-3	Sequence 3, Appl
27	40	32.8	512	1	US-07-623-953-5	Sequence 5, Appl
28	40	32.8	375	1	US-08-121-714-2	Sequence 2, Appl
29	40	32.8	483	1	US-08-468-700-34	Sequence 34, Appl
30	40	32.8	511	1	US-08-468-700-35	Sequence 35, Appl
31	40	32.8	1780	1	US-08-769-309A-5	Sequence 5, Appl
32	40	32.8	512	2	US-08-720-899-2	Sequence 2, Appl
33	40	32.8	483	2	US-08-645-971-2	Sequence 2, Appl
34	40	32.8	511	2	US-08-645-971-3	Sequence 3, Appl
35	40	32.8	375	2	US-08-477-108A-2	Sequence 2, Appl
36	40	32.8	512	2	US-08-459-610-2	Sequence 2, Appl
37	40	32.8	592	2	US-08-599-171A-30	Sequence 30, Appl
38	40	32.8	483	2	US-08-468-220-32	Sequence 32, Appl
39	40	32.8	511	2	US-08-468-220-33	Sequence 33, Appl

ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS, P.C.
STREET: 2026 Rambling Road
CITY: Kalamazoo
STATE: Michigan
COUNTRY: USA
ZIP: 49008-1699
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
COMPUTER: IBM PC/XT/AT Compatible
OPERATING SYSTEM: MS-DOS 5.0
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/717,365
FILING DATE: 23-SEP-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/528 255
FILING DATE: September 14, 1995
APPLICATION NUMBER: JP6-227876
FILING DATE: 22-SEPTEMBER-1994
APPLICATION NUMBER: JP7-183763
FILING DATE: 20-JULY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Terriyence F. Chapman
REGISTRATION NUMBER: 32549
REFERENCE/DOCKET NUMBER: Furuya Case 1335
TELECOMMUNICATION INFORMATION:
TELEPHONE: (616) 381-1156
TELEFAX: (616) 381-5465
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 482
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
LIBRARY: human fetal lung CDNA library
US-08-717-365-1

Query Match 35.2%; Score 43; DB 2; Length 482;
Best Local Similarity 40.0%; Pred. No. 38;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 6 TSGSTSDSKPOAETLKHKV 25
|| | : || : | : |
Db 460 TEEKTEKPEAKGVKEV 479

RESULT 6
US-08-313-553-15
Sequence 15, Application US/08313553
Patent No. 5641650
GENERAL INFORMATION:
APPLICANT: TURNER, George J.
APPLICANT: BETLACH, Mary C.
TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS POLYPEPTIDES
TITLE OF INVENTION: IN HALOBACTERIA
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Walter H. Dreger
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,553
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/038,662
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Walter H.
REGISTRATION NUMBER: 24,190
REFERENCE/DOCKET NUMBER: A-57669/WHd
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 559 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-313-553-15

Query Match 34.4%; Score 42; DB 1; Length 559;
Best Local Similarity 66.7%; Pred. No. 65;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 14 SKPQATLKHKV 25
: || | |||||
Db 280 ANPOPELLKHKV 291

RESULT 7
US-08-488-940-1
Sequence 1, Application US/08488940
Patent No. 5854049
GENERAL INFORMATION:
APPLICANT: Reed, Guy L.
TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,940
FILING DATE: 09-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/009001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1194 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-488-940-1

12

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,599
FILING DATE: 05-DEC-1995
CLASSIFICATION: 435

Query Match 33.6%; Score 41; DB 1; Length 830;

Search completed: November 13, 1999, 10:56:27
Job time: 1356 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:31 ; Search time 251.81 Seconds
(without alignments)
6.285 Million cell updates/sec

Title: US-08-913-430-3

Perfect score: 122

Sequence: 1 AGXGOTESGSDSKPQAEHLKHV 25

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA:*

1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep:*

2: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*

3: /cgn2_6/ptodata/1/paa/US080_COMB.pep:*

4: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*

5: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*

6: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*

7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep:*

8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep:*

9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*

10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*

11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*

12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*

13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*

14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*

15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*

16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*

17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*

18: /cgn2_6/ptodata/1/paa/US094_COMB.pep:*

19: /cgn2_6/ptodata/1/paa/US094A_COMB.pep:*

20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*

21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*

22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*

23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*

24: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	120	98.4	25	US-08-913-430-3
2	119	97.5	419	US-08-183-774-1
3	119	97.5	419	US-08-913-430-2
4	60	49.2	261	US-09-071-035-46
5	52	42.6	240	US-09-071-035-48
6	48	39.3	400	US-09-248-796-18257
7	48	39.3	400	US-08-096-409-18257
8	46	37.7	139	US-08-095-827-10
9	46	37.7	232	US-09-417-507-42910
10	45	36.9	1711	US-08-369-822A-10
11	45	36.9	1711	US-08-369-822A-10
12	45	36.9	1711	US-08-369-822C-10
13	45	36.9	1711	US-08-434-831B-10
14	45	36.9	1711	US-08-582-776A-10
15	45	36.9	1711	US-08-582-776C-10
16	45	36.9	621	US-09-026-343-7
17	45	36.9	186	US-09-270-767-46614
18	45	36.9	621	US-08-038-447-7
19	45	36.9	621	US-09-362-871-7

20	44	36.1	543	14	US-09-012-335-4	Sequence 4, Appli
21	44	36.1	433	14	US-09-071-035-314	Sequence 314, App
22	44	36.1	409	14	US-09-071-035-316	Sequence 316, App
23	44	36.1	254	15	US-09-134-000-4199	Sequence 4199, Ap
24	44	36.1	441	15	US-09-134-000-6501	Sequence 6501, Ap
25	44	36.1	1164	16	US-09-206-898-1	Sequence 1, Appli
26	44	36.1	433	16	US-09-270-767-31801	Sequence 31801, A
27	44	36.1	313	16	US-09-270-767-44375	Sequence 44375, A
28	44	36.1	433	16	US-09-270-767-47018	Sequence 47018, A
29	44	36.1	480	19	US-09-098-206-1	Sequence 1, Appli
30	44	36.1	754	19	US-09-117-905-1	Sequence 1, Appli
31	44	36.1	704	19	US-09-145-134-329	Sequence 329, App
32	44	36.1	1164	23	US-09-206-800-1	Sequence 1, Appli
33	43	35.2	482	10	US-08-624-735B-5	Sequence 5, Appli
34	43	35.2	199	11	US-08-737-248-12	Sequence 12, Appli
35	43	35.2	1097	13	US-08-951-188-4	Sequence 4, Appli
36	43	35.2	342	14	US-09-071-035-182	Sequence 182, App
37	43	35.2	322	14	US-09-071-035-184	Sequence 184, App
38	43	35.2	368	15	US-09-134-000-4284	Sequence 4284, Ap
39	43	35.2	587	16	US-09-270-767-39651	Sequence 39651, A
40	43	35.2	587	16	US-09-270-767-54868	Sequence 54868, A
41	43	35.2	485	17	US-09-328-352-8210	Sequence 8210, Ap
42	43	35.2	170	19	US-09-146-055-718	Sequence 718, App
43	43	35.2	170	24	US-09-147-189-761	Sequence 761, App
44	42	34.4	559	1	PCT-US94-02388-15	Sequence 15, Appli
45	42	34.4	263	19	US-09-096-409-20664	Sequence 20664, A

ALIGNMENTS

RESULT 1

US-08-913-430-3

Sequence 3, Application US/08913430B

GENERAL INFORMATION:

APPLICANT: Walker, John

APPLICANT: Lee, Rogan

TITLE OF INVENTION: Antigen Composition Against Mycoplasma

FILE REFERENCE: U-011415-0

CURRENT APPLICATION NUMBER: US/08/913,430B

CURRENT FILING DATE: 1997-12-09

EARLIER APPLICATION NUMBER: PCT/AU96/00149

EARLIER FILING DATE: 1996-03-15

EARLIER APPLICATION NUMBER: PN 1789

EARLIER FILING DATE: 1995-03-16

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Patentin Ver. 2.0 - beta

SEQ ID NO 3

LENGTH: 25

TYPE: PRT

ORGANISM: Mycoplasma hyopneumoniae

FEATURE:

NAME/KEY: Unsure

LOCATION: (3)

OTHER INFORMATION: Undetermined

US-08-913-430-3

Query Match 98.4%; Score 120; DB 13; Length 25;
Best Local Similarity 100.0%; Pred. No. 8.5e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGXGOTESGSDSKPQAEHLKHV 25
|||||

Db 1 AGXGOTESGSDSKPQAEHLKHV 25

RESULT 2

US-08-183-774-1

Sequence 1, Application US/08183774

GENERAL INFORMATION:

APPLICANT: SETO, Yasuhiro

```

; APPLICANT: FUTO, Satoshi
; APPLICANT: MITSUSE, Shizuo
; APPLICANT: MATSUO, Kanako
; APPLICANT: TSUNA, Mika
; TITLE OF INVENTION: DNA's Encoding Surface Antigen of
; TITLE OF INVENTION: Mycoplasma hyopneumoniae, DNA Fragments for Primer,
; TITLE OF INVENTION: Recombinant Antigenic Peptides and Diagnostic Method of
; TITLE OF INVENTION: Mycoplasma pneumoniae of Swine Using Same
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/183,774
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/747,015
; FILING DATE: 19-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, Norman F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4183-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 419 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-183-774-1

Query Match 97.5%; Score 119; DB 4; Length 419;
Best Local Similarity 96.0%; Pred. No. 3.1e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGXGQTESGSDSKPQAEHLKHV 25
Db 29 AGCGQTESGSDSKPQAEHLKHV 53

RESULT 3
US-08-913-430-2
; Sequence 2, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 2
; LENGTH: 419

```

```

; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
; US-08-913-430-2

Query Match 97.5%; Score 119; DB 13; Length 419;
Best Local Similarity 96.0%; Pred. No. 3.1e-10;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGXGQTESGSDSKPQAEHLKHV 25
Db 29 AGCGQTESGSDSKPQAEHLKHV 53

RESULT 4
US-09-071-035-46
; Sequence 46, Application US/09071035
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369p2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 261 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-46

Query Match 49.2%; Score 60; DB 14; Length 261;
Best Local Similarity 54.5%; Pred. No. 0.22;
Matches 12; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGXGQTESGSDSKPQAEHLK 22
Db 20 AGCGNKKTASTNDSKPQETK 41

RESULT 5
US-09-071-035-48
; Sequence 48, Application US/09071035
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496

```

RESULT 7

CC BY-NC-ND 4.0

RESULT 7

; NUMBER OF SEQUENCES:

```

RESULT 13
US-08-434-831B-10
US-08-434-831B-10 Application US/08434831B
GENERAL INFORMATION:
APPLICANT: Lipkin, W. I.
APPLICANT: Briese, Thomas
APPLICANT: Kliche, Stefanie
APPLICANT: Schneider, Patrick A.
APPLICANT: Stitz, Lothar
APPLICANT: Schneemann, Anette
TITLE OF INVENTION: Borna Disease Viral Sequences,
TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
TITLE OF INVENTION: System Diseases
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski, L.L.P.
STREET: 865 South Figueroa Street, 29th Floor
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017-2571
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0
SOFTWARE: ASCII (DOS) TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,831B
FILING DATE: 04-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/582,776A
FILING DATE: 04-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/369,822
APPLICATION NUMBER: APPLICATION NUMBER: US 08
FILING DATE: 06-JAN-1995
FILING DATE: FILING DATE: 04-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wong, Wean K.
REGISTRATION NUMBER: 33,561
REFERENCE/DOCKET NUMBER: 1279-194C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213/977-1001
TELEFAX: 213/977-1003
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:

Query Match 36.9%; Score 45; DB 9; Length 1711;
Best Local Similarity 47.6%; Pred. No. 3.9e+02;
Matches 10; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

Result No.	Score	Query Match	Length	DB	ID	Description
1	119	97.5	419	2	A56153	46K surface antigen
2	52	42.6	692	2	T03377	homeotic protein H
3	49	40.2	719	2	S25237	homeotic protein H
4	47	38.5	505	2	S4550	protein-tyrosine k
5	47	38.5	506	2	S4553	protein-tyrosine k
6	47	38.5	688	2	S3605	maturase-related p
7	47	38.5	4861	2	S71752	giant protein p619
8	46	37.7	414	2	A39205	nuclear localizati
9	46	37.7	320	2	JC4010	nucleotide-binding
10	45	36.9	194	2	S64075	probable ribosomal
11	45	36.9	512	2	B70785	probable integral
12	45	36.9	687	2	A34879	maturase-related p
13	45	36.9	621	2	S38880	eleven-nineteen ly
14	44	36.1	680	2	S09386	IGA-specific metal
15	44	36.1	1532	2	A36039	IGA-specific metal
16	44	36.1	218	2	T02912	probable transcrip
17	44	36.1	790	2	JC5749	DNA topoisomerase
18	44	36.1	88	2	S41030	hypothetical prote
19	44	36.1	543	2	A41211	early growth respo
20	44	36.1	220	2	A45597	acidic calmodulin-
21	43.5	35.7	714	2	S76082	hypothetical prote
22	43.5	35.7	924	2	S34926	hypothetical prote
23	43	35.2	229	1	LC9F	prolactin precursor
24	43	35.2	168	1	A37780	pleiotrophin precu
25	43	35.2	323	2	I49529	transcription fact
26	43	35.2	591	2	D64204	membrane lipoprote
27	43	35.2	399	2	JC5279	translocation prot
28	42.5	34.8	3924	2	S74331	ankyrin 2, neuro
29	42.5	34.8	235	2	S70219	sipa protein - Sal
30	42	34.4	311	1	DTECC	aspartate carbamoy
31	42	34.4	311	1	OWEBC	aspartate carbamoy
32	42	34.4	306	1	OWSEAC	aspartate carbamoy
33	42	34.4	483	2	A55033	keratin 12 - mouse
34	42	34.4	687	2	JN0667	heat-shock cognate
35	42	34.4	543	2	S46713	hypothetical prote
36	42	34.4	172	2	G99432	IGA-specific metal
37	42	34.4	370	2	S31831	IGA-specific metal
38	42	34.4	370	2	S40431	IGA-specific metal
39	42	34.4	324	2	S40432	IGA-specific metal

RESULT 3
S25237
homeotic protein Hoxla - maize

C:Species: Zea mays (maize)
C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 05-Dec-1997
C:Accession: S25237
R:Belmann, R.; Werr, W.
EMBO J. 11, 3367-3374, 1992
A:Title: Zmhoxia, the product of a novel maize homeobox gene, interacts with the Shrunken1 gene
A:Reference number: S25237; MUID:92371444
A:Accession: S25237
A:Molecule type: mRNA
A:Residues: 1-719 <BEL>
A:Cross-references: EMBL:X67561; NID:q22330; PID:q22331
C:Genetics:
A:Gene: hoxla
A:Map position: 8
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:560-616/Domain: homeobox homology <Hox>

Query Match 40.2%; Score 49; DB 2; Length 719;
Best Local Similarity 41.7%; Pred. No. 11;
Matches 10; Conservative 6; Mismatches 6; Indels 2; Gaps 1;

QY 2 GXGQTESGTSDSKPOAETLKHV 25
| : : : : : | : : : : : |
Db 12 GNKEIENGASSQNP--ESLEHSV 33

RESULT 4
S24550
Protein-tyrosine kinase (EC 2.7.1.112) 1 - freshwater sponge (Spongilla lacustris)
N:Alternate names: src-type tyrosine kinase 1
C:Species: Spongilla lacustris
C:Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 08-Sep-1997
C:Accession: S24550
R:Rauff, F.
Submitted to the EMBL Data Library, September 1991
A:Reference number: S24550
A:Accession: S24550
A:Molecule type: mRNA
A:Residues: 1-505 <RAU>
A:Cross-references: EMBL:X61601; NID:g10149; PID:g10150
C:Genetics:
A:Gene: srk1
C:Superfamily: protein-tyrosine kinase src; protein kinase homology; SH3 homology; SH3 homology
C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
F:61-111/Domain: SH3 homology <SH3>
F:122-214/Domain: SH2 homology <SH2>
F:238-496/Domain: protein kinase homology <KIN>
F:246-254/Region: protein kinase ATP-binding motif
F:268/Active site: Lys #status predicted

Query Match 38.5%; Score 47; DB 2; Length 505;
Best Local Similarity 37.5%; Pred. No. 15;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 2 GXGQTESGTSDSKPOAETLKHV 25
| : : : : : | : : : : : |
Db 12 GNKATAGSTVDSHLSQSVKGKI 35

RESULT 5
S24553
protein-tyrosine kinase (EC 2.7.1.112) 4 - freshwater sponge (Spongilla lacustris)
N:Alternate names: src-type tyrosine kinase 4
C:Species: Spongilla lacustris
C:Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 08-Sep-1997
C:Accession: S24553
R:Rauff, F.
Submitted to the EMBL Data Library, September 1991
A:Reference number: S24550
A:Accession: S24553

A:Molecule type: mRNA
A:Residues: 1-506 <RAU>
A:Cross-references: EMBL:X61604; NID:g10155; PID:g10156
C:Genetics:
A:Gene: srk4
C:Superfamily: protein-tyrosine kinase src; protein kinase homology; SH2 homology; SH2 homology; SH2 homology
C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
F:61-111/Domain: SH3 homology <SH3>
F:122-214/Domain: SH2 homology <SH2>
F:238-496/Domain: protein kinase homology <KIN>
F:246-254/Region: protein kinase ATP-binding motif
F:268/Active site: Lys #status predicted

Query Match 38.5%; Score 47; DB 2; Length 506;
Best Local Similarity 37.5%; Pred. No. 15;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 2 GXGQTESGTSDSKPOAETLKHV 25
| : : : : : | : : : : : |
Db 12 GNKATAGSTVDSHLSQSVKGKI 35

RESULT 6
S53605
maturase-related protein (nad1 intron) - soybean mitochondrion (fragment)
C:Species: mitochondrion Glycine max (soybean)
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 14-Aug-1998
C:Accession: S53605
R:Thomson, M.C.; Macfarlane, J.L.; Beagley, C.T.; Wolstenholme, D.R.
Nucleic Acids Res. 22, 5745-5752, 1994
A:Title: RNA editing of mat-r transcripts in maize and soybean increases similarity of protein.
A:Reference number: S53604; MUID:95140641
A:Accession: S53605
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-688 <THO>
A:Cross-references: GB:U09988; NID:g607800; PID:g840924
C:Genetics:
A:Gene: mitochondrion
C:Keywords: mitochondrion; RNA editing

Query Match 38.5%; Score 47; DB 2; Length 688;
Best Local Similarity 55.6%; Pred. No. 21;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 GQTESGTSDSKPOAETL 21
| : : : : : | : : : : : |
Db 434 GMSETGSLDGVPLAETL 451

RESULT 7
S71752
giant protein p619 - human
N:Alternate names: chromosome condensation regulator RCC1 homolog p619
C:Species: Homo sapiens (man)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 05-Jun-1998
C:Accession: S71752
R:Rosa, J.L.; Casaroli-Marano, R.P.; Buckler, A.J.; Vilaro, S.; Barbacid, M.
EMBO J. 15, 4262-4273, 1996
A:Title: p619, a giant protein related to the chromosome condensation regulator RCC1,
A:Reference number: S71752; MUID:97015127
A:Accession: S71752
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-4861 <ROS>
A:Cross-references: EMBL:U50078; NID:g1477564; PID:g1477565
C:Genetics:
A:Gene: p619
C:Function:
A:Description: may play an important role in the regulation of membrane trafficking;

C; Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat-containing proteins; RNA binding
C; Keywords: DNA binding; nucleus; RNA binding

Qy

```

: | : | | : : | | : | :
106 SSTCEAGSGAEEEEAKPEAKTV 126
Db

```

RESULT 11

B70786
probable integral membrane protein - Mycobacterium tuberculosis (strain H37RV)
C.Species: Mycobacterium tuberculosis
C.Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C.Accession: B70786
A:R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A.Accession: B70786
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-512 <COL>
A:Cross-references: GB:270283; GB:A123456; NID:g3261561; PID:g23585; PID:g1237062
A:Experimental source: strain H37RV
C:Genetics:
C:Gene: Rv2209

```
Query Match      36.9%; Score 45; DB 2; Length 512;
Best Local Similarity 52.9%; Pred. No. 31;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```

Qy 6 T E S G S T S D S K P Q A E T L K 22
 | | | | : | | | | |
Db 410 T R I G V T A D T S P O A A T L O 426

RESULT 12

A34879
maturase-related protein (nadl intron) - fava bean mitochondrion
C:Species: mitochondrion Vicia faba (fava bean)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 30-Sep-1993
C:Accession: A34879
R:Wahleithner, J.A.; Macfarlane, J.L.; Wolstenholme, D.R.
Proc. Natl. Acad. Sci. U.S.A. 87, 548-552, 1990
A:Title: A sequence encoding a maturase-related protein in a group II intron of a plant
A:Reference number: A34879; MUID:90138867
A:Accession: A34879
A:Molecule type: DNA
A:Residues: 1-687 <WAH>
A:Cross-references: EMBL:M30176
A:Note: the authors translated the codon CGG for residues 51, 210, 367, 378, 398, 485, 5
C:Genetics:
A:Genome: Mitochondrion
C:Keywords: mitochondrion

```
Query Match      36.9%; Score 45; DB 2; Length 687;
Best Local Similarity 55.6%; Pred. No. 42;
Matches 10: Conservative 1; Mismatches 7; Indels
```

Qy 4 QOTESGSTSDSKPQATL 21
| : | | | | | | |
Dp 433 GMSERGSLLDGVPIAETL 450

RESULT 13

I38880
 eleven-nineteen lysine-rich leukemia gene (ELL) protein - human
 C-Species: Homo-Sapiens (man)
 C-Accession: O1-Mar-1996 #sequence-revision 01-Mar-1996 #text_change 06-Sep-1996
 C-Date: O1-Mar-1996
 C-Accession: I38880
 R:ThyMan, M.J.: Levitan, P.A.; Kobayashi, H.; Simon, M.C.; Rowley, J.D.
 Proc. Natl. Acad. Sci. U.S.A. 91, 12110-12114, 1994

A;Title: Cloning of ELL, a gene that fuses to MLL in a t(11;19)(q23;p13.1) in acute myeloid leukemia
A;Reference number: I38880; MUID:95083651
A;Accession: I38880.
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-621 <RES>
A;Cross-references: EMBL:U16282; NID:g601792; PID:g601793
A;Experimental source: fetal brain CDNA library (commercially prepared)
C;Genetics:
A;Gene: GDB:ELL
A;Cross-references: GDB:450204
A;Map position: 19p13.1-19p13.1

Query Match 36.9%; Score 45; DB 2; Length 621;
Best Local Similarity 56.2%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 9 GSTSDSKPQAETLKHK 24
||| |||: :: |||
Db 441 GSPSRSKPKKSKKHK 450

RESULT
S09386

IGA-specific metalloendopeptidase (EC 3.4.24.13) - *Neisseria gonorrhoeae* (strain MS11)
 C.Species: *Neisseria gonorrhoeae*
 A.Variety: Strain MS11
 C.Date: 19-Mar-1997 #sequence_revision 24-Jul-1997 #text_change 21-Nov-1998
 C.Accession: S09386
 R.Halter, R.; Pohlner, J.; Meyer, T.F.
 EMBO J. 8, 2737-2744, 1989
 A.Title: Mosaic-like organization of IGA protease genes in *Neisseria gonorrhoeae* gene
 A.Reference number: S09386; MUID:90060036

A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-680 <HAL>
 A:Experimental source: strain MS11
 C:Genetics:
 A:Gene: iga
 C:Keywords: hydrolase; metalloproteinase; transmembrane protein
 F:410-411/Cleavage site: Pro-Ser (autolytic) #status predicted
 F:442-443/Cleavage site: Pro-Ser (autolytic) #status predicted
 F:545-546/Cleavage site: Pro-Ala (autolytic) #status predicted

Query Match 36.1%; Score 44; DB 2; Length 680;
Best Local Similarity 52.6%; Pred. No. 58;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 6 T E S G S T S D S K P Q A E T L K H K 24
: | | | | | | | | :
D6 625 S E S V S T S D K H P O D N T E I H E 643

RESULT 15
A26039

A26039 Iga-specific metalloendopeptidase (EC 3.4.24.13) precursor - *Neisseria gonorrhoeae*
 N;Alternate names: Iga protease; immunoglobulin A1 proteinase
 C;Species: *Neisseria gonorrhoeae*
 C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 20-Mar-1998
 C;Accession: A26039
 R;Polhner, J.; Halter, R.; Beyreuther, K.; Meyer, T.F.
 Nature 325, 458-462, 1987
 A;Title: Gene structure and extracellular secretion of *Neisseria gonorrhoeae* Iga, prot
 A;Reference number: A26039; PMID:87115823

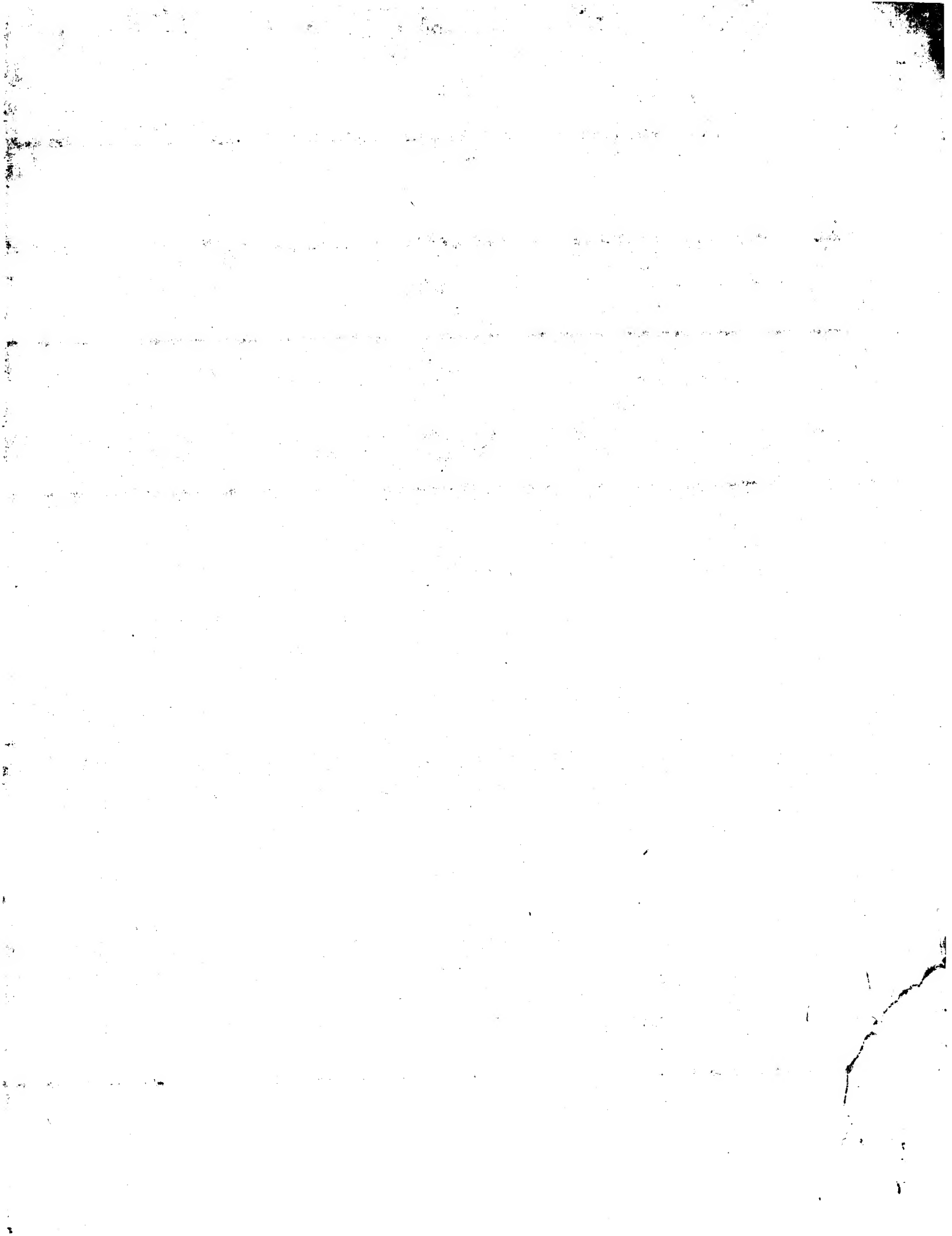
A;Molecule type: DNA
A;Residues: 1-1532 <POH>
A;Cross-references: GB:X04835; NID:g44868; PID:g44869
A;Note: the authors translated the codon AAG for resl
C;Keywords: hydrolase; metalloproteinase

F:1-27/Donaili; signal sequence #status predicted <SIG>
 F:28-1532/Product: immunoglobulin A1 proteinase #status predicted <MAT>

Query Match 36.18; Score 44; DB 2; Length 1532;
 Best Local Similarity 52.6%; Pred. NO. 1.4e+02;
 Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 6 TEGSTSDSKPQAE TLKHK 24
 Db 1201 SESVSTDKHPQDNTLHE 1219

Search completed: November 13, 1999, 12:07:52
 Job time: 2064 sec



Result No.	Score	Query		DB	ID	Description	
		Match	Length				
1	119	97.5	416	1	P46_MYCHY	P46192	mycoplasma
2	50	41.0	378	1	T54_HUMAN	Q92917	homo sapien
3	49	40.2	719	1	HX1A_WAIZE	P46605	zea mays (m
4	47	38.5	505	1	SRK1_SPOLA	P42686	spongilla 1
5	47	38.5	506	1	SRK4_SPOLA	P42690	spongilla 1
6	47	38.5	521	1	YAV8_SCHPO	Q10177	schizosacch
7	46	37.7	320	1	NBP_HUMAN	P53384	homo sapien
8	46	37.7	414	1	NSR1_YEAST	P27476	saccharomyc
9	45	36.9	621	1	ELL_HUMAN	P55199	homo sapien
10	45	36.9	1608	1	RRLP_BDV	P54639	borna disea
11	45	36.9	512	1	Y00K_MYCTU	Q10398	mycobacteri
12	45	36.9	194	1	YGG8_YEAST	P53163	saccharomyc
13	44	36.1	543	1	EGR1_HUMAN	P18146	homo sapien
14	44	36.1	1532	1	IGA_NEIGO	P09790	neisseria g
15	44	36.1	755	1	LEU2_RHIPU	P55251	rhizomucor
16	44	36.1	219	1	NAF2_RAT	Q05175	ratius norv
17	44	36.1	435	1	RA52_KLULA	P41768	kluyveromyc
18	43.5	35.7	924	1	YB3_YEAST	P38308	saccharomyc
19	43	35.2	482	1	HDA1_HUMAN	Q13547	homo sapien
20	43	35.2	482	1	HDA1_MOUSE	O09106	mus musculu
21	43	35.2	323	1	MAF1_MOUSE	P54841	mus musculu
22	43	35.2	323	1	MAF1_RAT	P54842	ratius norv
23	43	35.2	229	1	PRL_PG	P01238	sus scrofa
24	43	35.2	168	1	PTN_BOVIN	P21782	bos taurus
25	43	35.2	591	1	Y040_MYCGE	P47286	mycoplasma
26	42.5	34.8	3924	1	ANKB_HUMAN	Q01484	homo sapien
27	42	34.4	687	1	HSTF_DROME	P29845	drosophila
28	42	34.4	483	1	K1CL_MOUSE	Q64291	mus musculu
29	42	34.4	228	1	PRL_MONDO	O62819	monodelphis
30	42	34.4	310	1	PYRB_SCOLI	P00479	escherichia
31	42	34.4	310	1	PYRB_SALTY	P08420	salmonella
32	42	34.4	503	1	PYRB_SERMA	P19910	serratia ma
33	42	34.4	543	1	RRP3_YEAST	P38712	saccharomyc
34	42	34.4	252	1	TRV1_DROME	P52905	drosophila
35	41.5	34.0	510	1	HG2_HALRO	Q02508	halocynthia
36	41.5	34.0	440	1	STRP_STREQ	P00779	streptococc
37	41	33.6	404	1	9612_LYCES	P24396	lycopersico
38	41	33.6	312	1	BLAB_STRCI	P33652	streptomyce
39	41	33.6	754	1	CAS4_CANFA	Q28247	canis famli
40	41	33.6	1685	1	CAS4_HUMAN	P29400	homo sapien
41	41	33.6	640	1	DEXT_ARTGO	P70744	arthrobacte
42	41	33.6	549	1	EPD1_CANMA	P56092	candida mal
43	41	33.6	308	1	HMB1_STRPU	P13545	strongyloce

RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U66359; G1663764; -
SQ SEQUENCE 378 AA; 40804 MW; 5C3DEDE74 CRC32;

Query Match 41.0%; Score 50; DB 1; Length 378;
Best Local Similarity 55.6%; Pred. No. 2.8;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Qy 4 GOTESGTSKPKQAETL 21
| | | | | :|:|:|:|
Db 135 GCTPSGEGADSEAPRAETV 152

RESULT 3
HXIA_MAIZE STANDARD; PRT; 719 AA.
AC P46605;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HOMEBOX PROTEIN HXIA.
GN HXIA.
OS ZEA MAYS (MAIZE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; ZEA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92371444.
RA BELLMANN R., WERR W.;
RT "Zmhoxla, the product of a novel maize homeobox gene, interacts with
RT the Shrunken 26 bp feedback control element.";
RL EMOB J. 11:3367-3374 (1992).
CC -!- FUNCTION: INTERACTS WITH THE SHRUNKEN (SHR) 26 BP FEEDBACK CONTROL
CC ELEMENT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN KERNELS, LEAVES AND SHOOTS BUT
CC NOT IN ROOTS.
CC -!- SIMILARITY: STRONG, TO A.THALIANA HAT3.1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X67561; G22331; -
DR MAIZEDB: 25707; -
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
DR PFAM; PF00046; homeobox; 1.
DR PFAM; PF00628; PHD; 1.
DR HSSP; P22808; 1VND.
DR TRANSFAC; T00922; -
KW HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION.
FT DOMAIN 195 464 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 212 265 CYS-RICH (PHD-FINGER).
FT DNA_BIND 559 618 HOMEBOX.
SQ SEQUENCE 719 AA; 79116 MW; AF81EAB0 CRC32;

Query Match 40.2%; Score 49; DB 1; Length 719;
Best Local Similarity 41.7%; Pred. No. 8;
Matches 10; Conservative 6; Mismatches 6; Indels 2; Gaps 1;
Qy 2 GXGQTESGTSKPKQAETLKHV 25
| | | | | :|:|:|:|
Db 12 GNGEIENGASSSQNP--ESLEHSV 33

RESULT 4
SRKL_SPOLA STANDARD; PRT; 505 AA.
AC P42686;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE SRK1 (EC 2.7.1.112).
GN SRK1.
OS SPONGILLA LACUSTRIS (FRESHWATER SPONGE).
OC EUKARYOTA; METAZOA; PORIFERA; DEMOSPONGIAE; CERACTINOMORPHA;
OC HAPLOSCLERIDA; SPONGILLIDAE; SPONGILLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92334872.
RA OTTILIE S., RAULF F., BARNEKOW A., HANNIG G., SCHARTL M.;
RT "Multiple src-related kinase genes, srk1-4, in the fresh water sponge
RT Spongilla lacustris".
RL ONCOGENE 7:1625-1630 (1992).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- ALTERNATIVE PRODUCTS: SRK1 AND SRK4 MAY ARISE BY ALTERNATIVE
CC SPLICING.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X61601; G10150; -
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; P00523; 2PTK.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 54 116 SH3.
FT DOMAIN 122 214 SH2.
FT DOMAIN 240 493 PROTEIN KINASE.
FT NP_BIND 246 254 ATP (BY SIMILARITY).
FT BINDING 268 268 ATP (BY SIMILARITY).
FT ACT_SITE 359 359 BY SIMILARITY.
SQ SEQUENCE 505 AA; 57693 MW; 3AE3DF34 CRC32;

Query Match 38.5%; Score 47; DB 1; Length 505;
Best Local Similarity 37.5%; Pred. No. 11;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Qy 2 GXGQTESGTSKPKQAETLKHV 25

RESULT 6
YAV8-SCHPO
CC -!- SIMILARITY: BELONGS TO THE MRP/NBP35 FAMILY OF ATP-BINDING
RL GENE 147:281-285(1994).
RT protein related to Escherichia coli MIND.;

CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U01833; G515644; --
 CC MM: 600280; --
 CC PROSITE; PS01215; MRP; 1.
 CC NP_BIND 62 69 ATP (POTENTIAL).
 CC SEQUENCE 320 AA; 34588 MW; AF9ED113 CRC32;
 CC -----

Query Match 37.7%; Score 46; DB 1; Length 320;
 Best Local Similarity 32.0%; Pred. No. 9.4;
 Matches 8; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 1 AGXGTEGSGTSDSKPQAEYLKHKV 25
 DB 33 SGAGATPDATAEIKERKMTVKHKI 57

RESULT 8

NSR1_YEAST
 ID NSR1_YEAST STANDARD; PRT; 414 AA.
 AC P27476;
 DT 01-AUG-1992 (REL. 23, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE NUCLEAR LOCALIZATION SEQUENCE BINDING PROTEIN (P67).
 GN NSR1 OR YGR159C OR G7001.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACCHAROMYCETALES;
 CC SACCHAROMYCETACEAE; SACCHAROMYCES.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 92355583.
 RA KONDO K., INOUE M.;
 RT "Yeast NSR1 protein that has structural similarity to mammalian
 RT nucleolin is involved in pre-rRNA processing."
 RL J. BIOL. CHEM. 267:16252-16258(1992).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 92355583.
 RA KONDO K., INOUE M.;
 RT "Yeast NSR1 protein that has structural similarity to mammalian
 RT nucleolin is involved in pre-rRNA processing."
 RL J. BIOL. CHEM. 267:16252-16258(1992).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 96158062.
 RA SKALA J., NAWROCKI A., GOFFEAU A.;
 RT "The sequence of a 27 kb segment on the right arm of chromosome VII
 RT from Saccharomyces cerevisiae reveals MOL1, NAT2, RPL30B, RSRI, CYS4,
 RT PEM1/CHO2, NSR1 genes and ten new open reading frames."
 RL YEAST 11:1421-1427(1995).
 [4]
 RP SEQUENCE OF 188-414 FROM N.A.
 RA RIEGER M., MUELLER-AUER S., BRUECKNER M., SCHAEFER M.;
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 [5]
 RP DNA-BINDING.
 RX MEDLINE; 95098604.
 RA LIN J.-J., ZAKIAN V.A.;
 RT "Isolation and characterization of two Saccharomyces cerevisiae genes

RT that encode proteins that bind to (TGI-3)n single strand telomeric
 RT DNA in vitro."
 RL NUCLEIC ACIDS RES. 22:4906-4913(1994).
 CC -1- FUNCTION: INVOLVED IN PRE-RRNA PROCESSING. SPECIFICALLY BINDS
 CC NUCLEAR LOCALIZATION SEQUENCES. CANDIDATE FOR A RECEPTOR AT THE
 CC NUCLEUS THAT MAY BE INVOLVED IN BOTH RNA AND PROTEIN TRANSPORT.
 CC BINDS TELOMERIC SEQUENCES OF THE TYPE (TG[1-3])N IN VITRO.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; POSSIBLY AT THE NUCLEOLUS.
 CC -1- INDUCTION: IN RESPONSE TO LOW TEMPERATURE (BY COLD-SHOCK).
 CC -1- SIMILARITY: BELONGS TO THE GAR FAMILY OF PROTEINS.
 CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RNP).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X57185; G4058; --
 CC EMBL: X85807; G1045265; --
 CC EMBL: Z72944; E243542; --
 CC EMBL: Z72946; E312887; --
 CC PIR: A39205; A39205.
 CC SGD: L0001278; NSR1.
 CC PROSITE: PS00030; RNP_1; 2.
 CC PFM: PF00076; RNP_2.
 CC DNA-BINDING; RNA-BINDING; NUCLEAR PROTEIN; REPEAT; RRNA PROCESSING.
 FT DOMAIN 29 136 SER/ASP/GLU-RICH.
 FT DOMAIN 170 175 RNA-BINDING (RNP2) (BY SIMILARITY).
 FT DOMAIN 209 216 RNA-BINDING (RNP1) (BY SIMILARITY).
 FT DOMAIN 269 274 RNA-BINDING (RNP2) (BY SIMILARITY).
 FT DOMAIN 308 315 RNA-BINDING (RNP1) (BY SIMILARITY).
 FT DOMAIN 366 384 RNA-BINDING RGG-BOX (BY SIMILARITY).
 SQ SEQUENCE 414 AA; 44535 MW; FD58C3D3 CRC32;

 Query Match 37.7%; Score 46; DB 1; Length 414;
 Best Local Similarity 45.0%; Pred. No. 12;
 Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
 QY 5 QTESGTSKPKQAEYLKHK 24
 DB 65 ESSSSSSDSEAEATKKEE 84

 RESULT 9
 ELL_HUMAN
 ID ELL_HUMAN STANDARD; PRT; 621 AA.
 AC P55199;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE RNA POLYMERASE II ELONGATION FACTOR ELL (ELEVEN-NINETEEN LYSINE-RICH
 DE LEUKEMIA PROTEIN).
 GN ELL.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FETAL BRAIN;
 RX MEDLINE; 95083651.
 RA THIRMAN M.J., LEVITAN D.A., KOBAYASHI H., SIMON M.C., ROWLEY J.D.;
 RT "Cloning of ELL, a gene that fuses to MLL in a t(11;19)(q23;p13.1) in
 RT acute myeloid leukemia."
 RL PROC. NATL. ACAD. SCI. U.S.A. 91:12110-12114(1994).
 [2]
 RP FUNCTION.
 RX MEDLINE; 96175588.
 RA SHILATIPARD A., LANE W.S., JACKSON K.W., CONAWAY R.C., CONAWAY J.W.;
 RT "An RNA polymerase II elongation factor encoded by the human ELL

RT gene.";
RL SCIENCE 271:1873-1876(1996).
CC -!- FUNCTION: ELONGATION FACTOR THAT CAN INCREASE THE CATALYTIC RATE
CC OF RNA POLYMERASE II TRANSCRIPTION BY SUPPRESSING TRANSIENT
CC PAUSING BY THE POLYMERASE AT MULTIPLE SITES ALONG THE DNA.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES TESTED. HIGHEST
CC LEVELS FOUND IN PLACENTA, SKELETAL MUSCLE, TESTIS AND PERIPHERAL
CC BLOOD LEUKOCYTES.
CC -!- DISEASE: INVOLVED IN A T(11;19)(Q23;P13.1) CHROMOSOMAL
CC TRANSLOCATION IN ACUTE LEUKEMIAS CAUSING FUSION TO THE TRITHORAX
CC (MLL OR HRX) GENE PRODUCT WHICH CONTAINS DNA-BINDING MOTIFS
CC RESULTING IN A ROGUE ACTIVATOR PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE ELL / OCCUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U16282; G601793; ..
CC MIM: 600284; ..
CC NUCLEAR PROTEIN; TRANSCRIPTION REGULATION; CHROMOSOMAL TRANSLOCATION;
CC PROTO-ONCOGENE.
CC SITE 46 46 MLL FUSION POINT (IN ACUTE MYELOID
CC LEUKEMIA PATIENT).
CC FT DOMAIN 445 459 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT SEQUENCE 621 AA; 68264 MW; 03254229 CRC32;
CC -----
CC Query Match 36.9%; Score 45; DB 1; Length 621;
CC Best Local Similarity 56.2%; Pred. No. 27;
CC Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
CC -----
CC QY 9 GSTSDSKPQAEHLKH 24
CC ||| ||| : |||
CC DB 441 GSPSRKPKKKKKHK 456
CC -----
CC RESULT 10
CC RRPL_BDV STANDARD; PRT; 1608 AA.
CC AC P52639;
CC DT 01-OCT-1996 (REL. 34, CREATED)
CC DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
CC DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
CC DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
CC {L PROTEIN} (P180).
CC GN L.
CC OS BORNA DISEASE VIRUS (BDV).
CC OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=V;
CC RX MEDLINE; 94240137.
CC RA BRIESE T., SCHNEEMANN A., LEWIS A.J., PARK Y.-S., KIM S.,
CC RA LUDWIG H., LIPKIN W.I.;
CC RT "Genomic organization of Borna disease virus.";
CC RL PROC. NATL. ACAD. SCI. U.S.A. 91:4362-4366(1994).
CC -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC POLYMERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U04608; G511673; ..
KW TRANSFERASE; RNA-DIRECTED RNA POLYMERASE.
SQ SEQUENCE 1608 AA; 180254 MW; 55385702 CRC32;

Query Match 36.9%; Score 45; DB 1; Length 1608;
Best Local Similarity 47.6%; Pred. No. 76;
Matches 10; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 7 ESGSTSDSKP--QAEHLKHV 25
DB 1014 EKGYSDDARPSIQGGTLTHRL 1034

RESULT 11
Y00K_MYCTU STANDARD; PRT; 512 AA.
ID Y00K_MYCTU
AC Q10398;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 53.6 KD PROTEIN CY190.20.
GN MTCY190.20
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H378V;
RA MORPHY L., HARRIS D., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 270283; E233585; ..
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 55 75 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 238 258 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 359 379 POTENTIAL.
FT TRANSMEM 386 406 POTENTIAL.
SQ SEQUENCE 512 AA; 53578 MW; E56E52F1 CRC32;

Query Match 36.9%; Score 45; DB 1; Length 512;
Best Local Similarity 52.9%; Pred. No. 22;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 6 TBSGSTDSDSKPQAEHLK 22
DB 410 TRIGVTADTSPQAAIQQ 426

RESULT 12
YGG8_YEAST STANDARD; PRT; 194 AA.
ID YGG8_YEAST
AC P53163;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PUTATIVE MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L7/L12 PRECURSOR.
GN YGLO68W.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA: FUNGI; ASCOMYCOTA: HEMIASCOMYCETES; SACCCHAROMYCETALES;
CC SACCCHAROMYCETACEAE; SACCCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE: 97435481.
RA RIEGER M., BRUECKNER M., SCHAEFER M., MUELLER-AUER S.;
RT "Sequence analysis of 203 kilobases from *Saccharomyces cerevisiae*
chromosome VII";
RL YEAST 13:1077-1090(1997).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z72591; E243289; .
DR PFAM: PF00542; Ribosomal_L12; 1.
DR HSP: P02392; ICTF
KW HYPOTHETICAL PROTEIN; RIBOSOMAL PROTEIN; MITOCHONDRION;
FT TRANSIT PEPTIDE.
FT CHAIN 1 ?
FT CHAIN 194 ? MITOCHONDRIAL 60S RIBOSOMAL PROTEIN
L7/L12 HOMOLOG.
FT SEQUENCE 194 AA; 20650 MW; A6C8FEBD CRC32;
SQ

Query Match 36.9%; Score 45; DB 1; Length 194;
Best Local Similarity 33.3%; Pred. No. 7.7;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGXQTESGSTSDSKPOAETL 21
: |:::|:|:
DB 106 SSTGEAGSGAEAEAKPEAKTV 126

RESULT 13
EGRL_HUMAN STANDARD; PRT; 543 AA.
AC P18146;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE EARLY GROWTH RESPONSE PROTEIN 1 (EGR-1) (KROX24) (TRANSCRIPTION FACTOR
ETR103) (ZINC FINGER PROTEIN 225) (AT225).
GN EGR1 OR ZNF225.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA: METAZOA: CHORDATA: VERTEBRATA: MAMMALIA; EUTHERIA;
CC PRIMATES: CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-FORESKIN;
RX MEDLINE: 90332455.
RA SUGGS S.V., KATZOWITZ J.L., TSAI-MORRIS C., SUKHATME V.P.;
RT "cDNA sequence of the human cellular early growth response gene
Egr-1";
RL NUCLEIC ACIDS RES. 18:4283-4283(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX SHIMIZU N., OHTA M., FUJIWARA C., SAGARA J., MOCHIZUKI N., ODA T.,
RA UTIYAMA H.;
RT "A gene coding for a zinc finger protein is induced during 12-O-
tetradecanoylphorbol-13-acetate-stimulated HL-60 cell

RT differentiation";
RL J. BIOCHEM. 111:272-277(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90239549.
RA WRIGHT J.J., GUNTER K.C., MITSUYA H., IRVING S.G., KELLY K.,
RA SIEBENLIST U.;
RT "Expression of a zinc finger gene in HTLV-I- and HTLV-II-transformed
cells";
RL SCIENCE 248:588-591(1990).
CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR. RECOGNIZE AND BINDS TO THE
DNA SEQUENCE 5'-CGCCCCGC-3' (EGR-SITE). ACTIVATE THE TRANSCRIPTION
OF TARGET GENES WHOSE PRODUCTS ARE REQUIRED FOR MITOGENESIS AND
DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- INDUCTION: BY GROWTH FACTORS.
CC -1- SIMILARITY: TO THE ZINC FINGERS OF THE GROWTH REGULATED EGR
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52541; G31130; .
DR EMBL: M62829; G182263; .
DR EMBL: M80583; .; NOT_ANNOTATED_CDS.
DR PIR: A41211; A41211.
DR MIM: 128990; .
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 3.
DR PFAM: PF00096; zf-C2H2; 3.
DR HSP: P08046; IALL.
DR TRANSFAC; T00241; .
KW TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; NUCLEAR PROTEIN;
REPEAT; ZINC-FINGER; METAL-BINDING.
FT DOMAIN 57 84 SER/GLY-RICH.
FT DOMAIN 338 418 ZINC-FINGERS.
FT ZN_FING 338 362 C2H2-TYPE.
FT ZN_FING 368 390 C2H2-TYPE.
FT ZN_FING 396 418 C2H2-TYPE.
SQ SEQUENCE 543 AA; 57506 MW; E34B7AFE CRC32;

Query Match 36.1%; Score 44; DB 1; Length 543;
Best Local Similarity 47.4%; Pred. No. 33;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 GXGQTESGSTSDSKPOAET 20
: |:::|:|:
DB 74 GGSNSSSSSSTFNPOADT 92

RESULT 14
IGA_NEIGO STANDARD; PRT; 1532 AA.
ID IGA_NEIGO
AC P09790;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE IGA-SPECIFIC SERINE ENDOPEPTIDASE PRECURSOR (EC 3.4.21.72) (IGA
PROTEASE).
GN IGA.
OS NEISSERIA GONORRHOEA.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIAE; NEISSERIA.
RN [1]
RP SEQUENCE FROM N.A.; AND PARTIAL SEQUENCE.
RC STRAIN-MS11;
RX MEDLINE: 87115823.
RA POHLNER J., HALTER R., BEYREUTHER K., MEYER T.F.;
RT "Gene structure and extracellular secretion of *Neisseria gonorrhoeae*

```

CC -1- PATHWAY: SECOND STEP IN LEUCINE BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D67033; GI000692; .
CC DR PROSITE; PS00450; ACONITASE_1; 1.
CC DR PROSITE; PS01244; ACONITASE_2; 1.
CC DR PFAM; PF00330; aconitase; 1.
CC DR PFAM; PF00694; Aconitase_C; 1.
CC KW LEUCINE BIOSYNTHESIS; LYASE; IRON-SULFUR; 4FE-4S. (BY SIMILARITY).
CC FT METAL 353 353 IRON-SULFUR (4FE-4S)
CC FT METAL 413 413 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC FT METAL 416 416 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC SQ SEQUENCE 755 AA; 82486 MW; 580EA467 CRC32;
CC -----
Query Match 36.1%; Score 44; DB 1; Length 755;
Best Local Similarity 57.1%; Pred. No. 47;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

Qy	12	SDSKFQAETLKHV	25
Db	83	ADSRTOCETLEHNV	96

Search completed: November 13, 1999, 10:33:29
Job time: 5179 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:20 ; Search time 139.86 Seconds
(without alignments)
11.001 Million cell updates/sec

Title: US-08-913-430-3

Perfect score: 122

Sequence: 1 ACXGQTESGSTSDSKPQAEITLKHV 25

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPREMBL10.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_prodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	42.6	692	10 024569	024569 zea mays (m
2	49	40.2	474	5 061472	061472 alysia cal
3	48	39.3	665	10 049236	049236 arena fatua
4	47	38.5	4861	4 015751	015751 homo sapien
5	47	38.5	688	8 034597	034597 glycine max
6	46	37.7	173	3 013320	013320 metarhizium
7	45	36.9	635	2 066065	066065 fibrobacter
8	45	36.9	1081	4 043485	043485 homo sapien
9	45	36.9	480	4 043499	043499 homo sapien
10	45	36.9	437	4 043500	043500 homo sapien
11	45	36.9	453	4 043501	043501 homo sapien
12	45	36.9	453	4 043492	043492 homo sapien
13	45	36.9	480	4 043493	043493 homo sapien
14	45	36.9	437	4 015282	015282 homo sapien
15	45	36.9	985	4 075184	075184 homo sapien
16	45	36.9	1099	4 060721	060721 homo sapien
17	45	36.9	3488	5 091257	091257 caenorhabdi
18	45	36.9	665	8 036216	036216 vicia faba
19	45	36.9	1534	12 088627	088627 borna disea
20	44	36.1	742	3 013853	013853 schizosacch
21	44	36.1	1150	3 079045	079045 emericeila
22	44	36.1	437	4 092760	092760 homo sapien
23	44	36.1	88	5 021196	021196 caenorhabdi
24	44	36.1	914	10 093659	093659 brassica na
25	44	36.1	988	10 093065	093065 brassica na
26	44	36.1	790	10 093119	093119 daucus caro
27	44	36.1	218	10 0927E3	0927E3 arabidopsis
28	43.5	35.7	714	2 055559	055559 synechocyst
29	43	35.2	399	4 099442	099442 homo sapien

ALIGNMENTS

RESULT 1

024569 PRELIMINARY; PRT; 692 AA.

AC 024569;
DT 01-JAN-1998 (TREMREL. 05, Created)
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
DE HOXB1 PROTEIN.
GN HOXB1.

OS zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eukaryotes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Zea.
RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 96361874.

RA UEBERLACKER B., KLINGE B., WERR W.;

RT "Ectopic expression of the maize homeobox genes ZmHox1a or ZmHox1b causes pleiotropic alterations in the vegetative and floral development of transgenic tobacco.";
RL Plant Cell 8:349-362(1996).

DR EMBL; X92428; CAA63156.1; -.

DR PFAM; PF00046; homeobox; 1.

DR PFAM; PF00628; PHD; 1.

KW Homeobox; Nuclear protein; Homeobox.

SQ SEQUENCE 692 AA; 76674 MW; 2053B9F6 CRC32;

Query Match 42.6%; Score 52; DB 10; Length 692;

Best Local Similarity 45.8%; Pred. No. 4.5;

Matches 11; Conservative 5; Mismatches 6; Indels 2; Gaps 1;

QY 2 GXGQTESGSTSDSKPQAEITLKHV 25

Db 12 NGGEITNGTSSQNP--ETLEHPV 33

RESULT 2

061472 PRELIMINARY; PRT; 474 AA.

AC 061472;

DT 01-AUG-1998 (TREMREL. 07, Created)

DT 01-AUG-1998 (TREMREL. 07, Last sequence update)

DT 01-AUG-1998 (TREMREL. 07, Last annotation update)

DE REDUCTASE-RELATED PROTEIN.

OS Aplysia californica (California sea hare).

OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidae;

OC Aplysiidae; Aplysia.

RN [1]

RP SEQUENCE FROM N.A.

RA ZWARTJES R.E., WEST H., HATTAR S., REN X., NOEL F., NUNEZ-REGUEIRO M.,

MACPHEE K., HOMAYOUNI R., CROW M.T., BYRNE J.H., ESKIN A.;

QY 7 ESGSTSDSKPOAETLKHV 25
 DB 143 KNGAPADKQPEAEGLKNOI 161

RESULT 7
 O66065 PRELIMINARY; PRT; 635 AA.

AC O66065;
 DT 01-AUG-1998 (TEMBLrel. 07, Created)
 DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
 DE CMC-XLANASE (FRAGMENT).
 OS Fibrobacter succinogenes S85.
 OC Bacteria; Fibrobacter group; Fibrobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S85;
 RA CHO K.K.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U94826; AAC06197.1; -
 DR PFAM; PF00553; CBD.2; 1.
 DR PFAM; PF00150; cellulase; 1.
 KW Xylan degradation; Hydrolase; Glycosidase.
 FT NON_TER 635
 SQ SEQUENCE 635 AA; 60943 MW; 2510AB57 CRC32;

Query Match 36.9%; Score 45; DB 2; Length 635;
 Best Local Similarity 50.0%; Pred. No. 48;
 Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 4 GQTEGSGTSDSKPOAETL 21
 DB 585 GQPSDGPSPDAKPAASAL 602

RESULT 8
 O43485 PRELIMINARY; PRT; 1081 AA.

AC O43485;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
 DT 01-JUN-1998 (TEMBLrel. 06, Last annotation update)
 DE RETINAL ROD NA+/CA+, K+ EXCHANGER.
 GN NCKX.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TUCKER J.E., WINKFEIN R.J., COOPER C.B., SCHNETKAMP P.P.M.;
 RL Invest. Ophthalmol. Vis. Sci. 39:435-440(1998).
 DR EMBL; AF026132; AAB97832.1; -
 SQ SEQUENCE 1081 AA; 119494 MW; 0243A458 CRC32;

Query Match 36.9%; Score 45; DB 4; Length 1081;
 Best Local Similarity 42.1%; Pred. No. 83;
 Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGXGQTEGSGTSDSKPOAE 19
 DB 746 AGEGETEKSGETQPEGE 764

RESULT 9
 O43499 PRELIMINARY; PRT; 480 AA.

AC O43499;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
 DT 01-JUN-1998 (TEMBLrel. 06, Last annotation update)

DE HTGN51.
 GN TGN.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KAIN R., ANGATA K., KERJASCHKI D., FUKUDA M.;
 RL J. Biol. Chem. 273:0-0(1998).
 DR EMBL; AF029316; AAB96908.1; -
 DR EMBL; AF029313; AAB96908.1; JOINED.
 DR EMBL; AF029314; AAB96908.1; JOINED.
 DR EMBL; AF029315; AAB96908.1; JOINED.
 SQ SEQUENCE 480 AA; 50992 MW; 9F25437D CRC32;

Query Match 36.9%; Score 45; DB 4; Length 480;
 Best Local Similarity 55.6%; Pred. No. 36;
 Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 5 QTEGSGTSDSKPOAETLK 22
 DB 103 QTPKGSTSKSGSEAQT 120

RESULT 10
 O43500 PRELIMINARY; PRT; 437 AA.

AC O43500;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
 DT 01-JUN-1998 (TEMBLrel. 06, Last annotation update)
 DE HTGN46.
 GN TGN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KAIN R., ANGATA K., KERJASCHKI D., FUKUDA M.;
 RL J. Biol. Chem. 273:0-0(1998).
 DR EMBL; AF029316; AAB96906.1; -
 DR EMBL; AF029313; AAB96906.1; JOINED.
 DR EMBL; AF029314; AAB96906.1; JOINED.
 DR EMBL; AF029315; AAB96906.1; JOINED.
 SQ SEQUENCE 437 AA; 45759 MW; A7885357 CRC32;

Query Match 36.9%; Score 45; DB 4; Length 437;
 Best Local Similarity 55.6%; Pred. No. 33;
 Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 5 QTEGSGTSDSKPOAETLK 22
 DB 103 QTPKGSTSKSGSEAQT 120

RESULT 11
 O43501 PRELIMINARY; PRT; 453 AA.

AC O43501;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
 DT 01-JUN-1998 (TEMBLrel. 06, Last annotation update)
 DE HTGN48.
 GN TGN.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KAIN R., ANGATA K., KERJASCHKI D., FUKUDA M.;
 RL J. Biol. Chem. 273:0-0(1998).

DR EMBL; AF029316; AAB96907.1; -.
DR EMBL; AF029313; AAB96907.1; JOINED.
DR EMBL; AF029314; AAB96907.1; JOINED.
DR EMBL; AF029315; AAB96907.1; JOINED.
SQ SEQUENCE 453 AA; 47577 MW; EE58FD18 CRC32;

Query Match 36.9%; Score 45; DB 4; Length 453;
Best Local Similarity 55.6%; Pred. No. 34;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 5 QTESGSTSDSKPQAEITLK 22
|| ||||| :|||
Db 103 QTPKGSTSKSGSEAQTTK 120

RESULT 12
O43492 PRELIMINARY; PRT; 453 AA.
AC O43492;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE HTGN48 PRECURSOR.
GN TGN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 98086273.
RA KAIN R., ANGATA K., KERJASCHKI D., FUKUDA M.;
RT "Molecular cloning and expression of a novel human trans-Golgi
network glycoprotein, TGN51, that contains multiple
tyrosine-containing motifs";
RL J. Biol. Chem. 273:981-988(1998).
DR EMBL; AF027515; AAC39541.1; -.
KW Signal.
FT SIGNAL
SQ SEQUENCE 453 AA; 47591 MW; A4542DF7 CRC32;

Query Match 36.9%; Score 45; DB 4; Length 453;
Best Local Similarity 55.6%; Pred. No. 34;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 5 QTESGSTSDSKPQAEITLK 22
|| ||||| :|||
Db 103 QTPKGSTSKSGSEAQTTK 120

RESULT 13
O43493 PRELIMINARY; PRT; 480 AA.
AC O43493;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE HTGN51 PRECURSOR.
GN TGN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 98086273.
RA KAIN R., ANGATA K., KERJASCHKI D., FUKUDA M.;
RT "Molecular cloning and expression of a novel human trans-Golgi
network glycoprotein, TGN51, that contains multiple
tyrosine-containing motifs";
RL J. Biol. Chem. 273:981-988(1998).

DR EMBL; AF027516; AAC39542.1; -.
KW Signal.
FT SIGNAL
SQ SEQUENCE 480 AA; 51006 MW; C91F59B1 CRC32;

Query Match 36.9%; Score 45; DB 4; Length 480;
Best Local Similarity 55.6%; Pred. No. 36;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 5 QTESGSTSDSKPQAEITLK 22
|| ||||| :|||
Db 103 QTPKGSTSKSGSEAQTTK 120

RESULT 14
O15282 PRELIMINARY; PRT; 437 AA.
AC O15282;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE TRANS-GOLGI 38 GLYCOPROTEIN.
GN HTG38.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE; 98086273.
RA KAIN R., ANGATA K., KERJASCHKI D., FUKUDA M.;
RT "Molecular cloning and expression of a novel human trans-Golgi
network glycoprotein, TGN51, that contains multiple
tyrosine-containing motifs";
RL J. Biol. Chem. 273:981-988(1998).
DR EMBL; U62390; AAC39539.1; -.
SQ SEQUENCE 437 AA; 45773 MW; 7CC86ED4 CRC32;

Query Match 36.9%; Score 45; DB 4; Length 437;
Best Local Similarity 55.6%; Pred. No. 33;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 5 QTESGSTSDSKPQAEITLK 22
|| ||||| :|||
Db 103 QTPKGSTSKSGSEAQTTK 120

RESULT 15
O75184 PRELIMINARY; PRT; 985 AA.
AC O75184;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE KIAA0702 PROTEIN.
GN KIAA0702.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 98403880.
RA ISHIKAWA K., NAGASE T., SUYAMA M., MIYAJIMA N., TANAKA A., KOTANI H.,
RA NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro";
RL DNA Res. 5:169-176(1998).
DR EMBL; AB014602; BAA31677.1; -.
SQ SEQUENCE 985 AA; 109391 MW; 1BFD4EFD CRC32;

Query Match 36.9%; Score 45; DB 4; Length 985;
Best Local Similarity 42.1%; Pred. NO. 75;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AGXGOTESGSTSDSKPQAE 19
||| |::| |
Db 746 AGEGETEERSGGTQPEGE 764

Search completed: November 13, 1999, 12:55:22
Job time: 3021 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:59:45 ; Search time 104.22 Seconds
(without alignments)
6.591 Million cell updates/sec

Title: US-08-913-430-4

Perfect score: 138

Sequence: 1 TIYPDKVLGVAVELRLIAKNKASR 29

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq.36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	138	100.0	281	1 R06279	Swine enzootic pne
2	138	100.0	29	1 W01034	Mycoplasma 46-48 k
3	138	100.0	419	1 W01037	Mycoplasma 46-48 k
4	131	94.9	419	1 R21829	Sequence of surfac
5	51	37.0	400	1 W88434	Disease associated
6	48	34.8	334	1 W06319	Human mitogen-acti
7	48	34.8	334	1 W19631	Human mitogen-acti
8	48	34.8	260	1 W97664	Human mitogen-acti
9	48	34.8	419	1 W97665	Human mitogen-acti
10	48	34.8	380	1 W97666	Mouse mitogen acti
11	48	34.8	453	1 W97667	Mouse mitogen acti
12	48	34.8	334	1 W97669	Human mitogen acti
13	48	34.8	389	1 W97663	Mouse mitogen acti
14	48	34.8	468	1 W94853	Mouse mutant MAPK
15	48	34.8	435	1 W94849	Human MAPK kinase
16	48	34.8	468	1 W94850	Mouse MAPK kinase
17	48	34.8	419	1 W94851	Mouse MAPK kinase
18	48	34.8	333	1 Y05335	Human SKK4 protein
19	47	34.1	318	1 W06318	Human mitogen-acti
20	47	34.1	318	1 W97668	Human mitogen acti
21	46	33.3	430	1 W20585	H. pylori secreted
22	46	33.3	393	1 W32867	Mitogen activated
23	46	33.3	429	1 Y10949	H. pylori ORF 04cp
24	44	31.9	146	1 W20201	H. pylori secreted
25	44	31.9	345	1 W40422	S. glaucosces acb
26	44	31.9	127	1 W27927	Amino acid sequenc
27	43.5	31.5	486	1 Y04660	L.lactis HsdM subu
28	43.5	31.5	515	1 Y04660	Yeast RNA-binding
29	43	31.2	397	1 P70261	Sequence of rat gl
30	43	31.2	346	1 W97662	Mouse mitogen acti
31	42	30.4	397	1 P80699	Deduced sequence o
32	42	30.4	397	1 P80700	Deduced sequence o
33	42	30.4	398	1 P70265	Sequence of human
34	42	30.4	397	1 R20087	PN-I alpha analogu
35	42	30.4	397	1 R20000	Sequence of protea
36	42	30.4	397	1 R20086	PN-I alpha analogu
37	42	30.4	397	1 R20088	PN-I alpha analogu
38	42	30.4	398	1 R21995	PN-I beta analogu
39	42	30.4	397	1 R20089	PN-I alpha analogu
40	42	30.4	398	1 R21998	PN-I beta analogu
41	42	30.4	398	1 R21996	PN-I beta analogu
42	42	30.4	398	1 R21997	PN-I beta analogu
43	42	30.4	397	1 R74557	Protease Nexin-1 t

44 42 30.4 398 1 R74558
45 42 30.4 344 1 W54370
Protease Nexin-1 t
G-protein coupled

ALIGNMENTS

RESULT 1

R06279
ID R06279 standard; protein; 281 AA.
AC R06279;
DT 19-DEC-1990 (first entry)
DE Swine enzootic pneumonia mycoplasma surface antigen polypeptide.
KW Mycoplasma pneumoniae; enzootic pneumonia; ds.
OS Mycoplasma hyopneumoniae.
PN J02167079-A.
PD 27-JUN-1990.
PF 21-DEC-1988; 322829.
PR 21-DEC-1988; JP-322829.
PA (NISE-) NIPPON SEIFUN KK.
DR WPI; 90-241949/32.
DR N-PSDB; Q05578.
PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
PT for early detection and treatment of mycoplasma or enzootic
PT pneumonia of pigs
PS Claim 1; Page 570; 28pp; Japanese.
CC Sequence encoding polypeptides may be used as probes for the early
CC diagnosis of pneumonia in pigs. The sequences may be used to produce
CC an expression vector and transform a suitable host.
SQ Sequence 281 AA;

Query Match 100.0%; Score 138; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 1.9e-13;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TIYPDKVLGVAVELRLIAKNKASR 29
|||||

DB 203 TIYPDKVLGVAVELRLIAKNKASR 231
|||||

RESULT 2

W01034
ID W01034 standard; Peptide; 29 AA.
AC W01034;
DT 19-JAN-1997 (first entry)
DE Mycoplasma 46-48 kDa protective antigen fragment CNBR Fl.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumonia;
OS Mycoplasma hyopneumoniae strain Beaufort.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149
PR 16-MAR-1995; AU-001789.
PA (UYME) UNIV MELBOURNE.
DR Doughty SW, Lee R, Walker J;
DR WPI; 96-433763/43.
PT Putative protective antigens against Mycoplasma - used for the
PT detection, prevention or treatment of Mycoplasma infections, esp. M.
PT hyopneumoniae in swine
PS Claim 13; Page 28; 43pp; English.
CC A 46-48 kDa putative protective antigen against Mycoplasma contains
CC the N-terminal sequence given in W01033 and the internal CNBR
CC fragments given in W01034-36. The antigen was isolated from
CC Mycoplasma hyopneumoniae cells using antibody probes enriched with
CC Mycoplasma-specific antibodies. Other protective antigens were
CC also identified (see also W01024-32). A gene (T38241) coding for
CC the 48 kDa antigen (W01037) was isolated from a genomic library.
CC Protective antigens and antibodies can be used in vaccines for
CC preventing or treating mycoplasma infections, partic. M.
CC hyopneumoniae infections in swine. They can also be used for
SQ diagnosis. 29 AA;
Sequence 29 AA;


```

Query Match          100.0%; Score 138; DB 1; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TIYPKPKVLGKVAVEVLRVLIKKKASR 29
   |||||
Db 1 TIYPKPKVLGKVAVEVLRVLIKKKASR 29

RESULT 3
W01037
ID W01037 standard; Protein: 419 AA.
AC W01037;
DE 19-JAN-1997 (first entry)
KW Mycoplasma 46-48 kDa protective antigen.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumonia;
KW diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
FH Key
FT misc_difference 70
FT /note= "codon 70 in the nucleotide sequence is
FT a stop codon"
FT misc_difference 101
FT /note= "codon 101 in the nucleotide sequence is
FT a stop codon"
FT misc_difference 254
FT /note= "codon 254 in the nucleotide sequence is
FT is a stop codon"
PN W09628472-A1.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (UYNE ) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
PI WPI; 96-433763/43.
DR N-PSDB; T38241.
PT Putative protective antigens against Mycoplasma - used for the
PT detection, prevention or treatment of Mycoplasma infections, esp. M.
PT hyopneumoniae in swine
PS Disclosure: Fig 7; 43pp; English.
CC The gene (T38241) coding for a 48 kDa putative protective antigen
CC (W01037) was obt'd. from a Mycoplasma hyopneumoniae genomic library
CC by screening with a probe generated by PCR amplification (see also
CC T13814-16). The antigen had originally been isolated from M.
CC hyopneumoniae cells using antibody probes enriched with Mycoplasma-
CC specific antibodies. Other protective antigens were also identified
CC (see also W01024-31). Protective antigens and antibodies can be
CC used in vaccines for preventing or treating mycoplasma infections,
CC partic. M. hyopneumoniae infections in swine. They can also be used
CC for diagnosis.
SQ Sequence 419 AA;

Query Match          100.0%; Score 138; DB 1; Length 419;
Best Local Similarity 100.0%; Pred. No. 2.9e-13;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TIYPKPKVLGKVAVEVLRVLIKKKASR 29
   |||||
Db 341 TIYPKPKVLGKVAVEVLRVLIKKKASR 369

RESULT 4
R21829
ID R21829 standard; Protein: 419 AA.
AC R21829;
DE 02-NOV-1992 (first entry)
KW Sequence of surface antigen 46kd.
KW Swine pneumonia; epidemic; diagnosis; therapy.
OS Mycoplasma hyopneumoniae.
PN EP-475185-A.

```

```

PD 18-MAR-1992.
PF 27-AUG-1991; 114335.
PR 27-AUG-1990; JP-224945.
PA (NIFL-) NIPPON FLOUR MILLS.
PI Seto Y, Futo S, Mitsuse S, Matsuo K, Tsuna M;
DR WPI; 92-089874/12.
DR N-PSDB; Q22042.
PT DNA and peptide of mycoplasma hyopneumoniae - useful for
PT diagnosis and treatment of swine mycoplasma pneumonia
PS Disclosure: Page 4-6 and pages 34-36; 45pp; English.
CC The inventors claim DNA encoding a surface antigen and primers used
CC in a method of diagnosing mycoplasma pneumonia of swine.
CC Mycoplasma cells were collected from the culture of M.hp, lysed in a
CC buffer containing SDS, followed by the extraction of DNA and
CC purification thereof. The DNA of M.hp is cleaved with a restriction
CC enzyme HindIII, the resulting fragment is inserted into a plasmid
CC pUC19 at the site cleaved with HindIII. This plasmid is transformed
CC into E. coli as a host cell, colony hybridization is performed by the
CC DNA probe method utilizing EcoRI fragments of pKUM1 and pKUM2 to
CC select the bacterial cells containing a plasmid (pURR126) carrying the
CC 46 kD antigenic gene. E. coli JM 109 which comprises the plasmid
CC pKUM1 or pKUM2 is deposited with FRI under accession No. FERM P-10318
CC or P-10319. The base sequence of the 46 kD antigen gene is given in
CC Q22042.
SQ Sequence 419 AA;

Query Match          94.9%; Score 131; DB 1; Length 419;
Best Local Similarity 96.6%; Pred. No. 3.3e-12;
Matches 28; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TIYPKPKVLGKVAVEVLRVLIKKKASR 29
   |||||
Db 341 TIYPKPKVLGKVAVEVLRVLIKKKASR 369

RESULT 5
W88434
ID W88434 standard; Protein: 400 AA.
AC W88434;
DE 26-APR-1999 (first entry)
DE Disease associated protein kinase DAPK-3.
KW DAPK-3; disease associated protein kinase; human; diagnosis;
KW therapy; adult respiratory distress syndrome; allergy; asthma;
KW arteriosclerosis; bronchitis; emphysema; hypereosinophilia;
KW myocardial inflammation; pericardial inflammation; anaemia;
KW rheumatoid arthritis; Addison's disease; AIDS; atherosclerosis;
KW atopic dermatitis; dermatomyositis; diabetes mellitus;
KW glomerulonephritis; gout; Grave's disease; lupus erythematosus;
KW multiple sclerosis; myasthenia gravis; osteoarthritis;
KW osteoporosis; pancreatitis; polycystic kidney disease;
KW polymyositis; scleroderma; Sjogren's syndrome;
KW autoimmune thyroiditis; cancer; infection; trauma;
KW cell proliferation.
OS Homo sapiens.
FH Key
FT Binding_site 79..86
FT /note= "potential ATP binding site"
FT Peptide 112..119
FT /note= "catalytic loop"
PN W09858052-A2.
PD 23-DEC-1998.
PF 19-JUN-1998; U12813.
PR 19-JUN-1997; US-878989.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Corley NC, Goli SK, Guegler KJ, Hillman JL,
PI Lal P, Shah P;
DR WPI; 99-080952/07.
DR N-PSDB; X06833.
PT New disease associated protein kinases - used to stimulate cell
PT proliferation and to treat the immune response and cancer
PS Claim 1; Page 57-58; 93pp; English.
CC This is the amino acid sequence of human disease associated protein

```

kinase DAPK-3, as deduced from a consensus sequence (see X06833) of overlapping cDNA clones from libraries which are immortalised or cancerous and show inflammatory or immune responses. DAPK-3 shows 94% homology to rat MAP kinase MEK2 (GI 1488263). The invention provides DAPK-1 to DAPK-7 polypeptides (see W88432-38) and cDNA clones encoding them (see X06831-36 and X06882), as well as expression vectors, host cells, agonists, antagonists and antibodies. The invention further provides uses of such products in the diagnosis, prevention and treatment of diseases associated with cell proliferation, especially cancer or an immune response (claimed). Conditions that may be treated include adult respiratory distress syndrome, allergies, asthma, arteriosclerosis, bronchitis, emphysema, hyperosinophilia, myocardial or pericardial inflammation, rheumatoid arthritis, Addison's disease, AIDS, anaemia, atherosclerosis, various diseases of the digestive system, atopic dermatitis, dermatomyositis, diabetes mellitus, glomerulonephritis, gout, Grave's disease, lupus erythematosus, multiple sclerosis, myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis, polycystic kidney disease, polymyositis, scleroderma, Sjorgren's syndrome, autoimmune thyroiditis, complications of cancer, extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal and helminthic infections, and trauma (disclosed).
Sequence 400 AA;

Query Match 37.0%; Score 51; DB 1; Length 400;
Best Local Similarity 45.5%; Pred. No. 3.5;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLIAKNNK 26
|:::||||: ||| | : |
DB 166 PEEILGKVSIAVLRLGLAYLRK 187

RESULT 6
W06319

ID W06319 standard; Protein; 334 AA.
AC W06319;
DT 07-FEB-1997 (first entry)
DE Human mitogen-activated protein kinase kinase 6.
KW Tyrosine kinase; signal transduction; cytokine; oncoprotein;
KW stress; diagnosis; therapy.
OS Homo sapiens.
PN W09636642-A1.
PD 21-NOV-1996; U01078.
PF 26-JAN-1996; U01078.
PR 19-MAY-1995; US-446083.
PR 19-SEP-1995; US-530950.
PA (DAVI/) DAVIS R J.
PA (DERI/) DERIJARD B.
PA (GUPT/) GUPTA S.
PA (RAIN/) RAINGEAUD J.
PI Davis RJ, Derijard B, Gupta S, Raingaud J;
DR WPI: 97-012035/01.
DR N-PSDB; T43204.
PT New mitogen activated protein kinase - useful for treating
PT ischemic heart disease, kidney failure etc., also for identifying
PT modulators for treatment of similar conditions
PS Claim 6; Fig 5; 104pp; English.
CC Novel human mitogen activated protein kinase kinase 6 (MKK6)
CC mediates a signal transduction pathway that activates human
CC mitogen-activated protein (MAP) kinase p38. Its amino acid
CC sequence was deduced from a cDNA clone (T43204) obtd. from a human
CC skeletal muscle library. Recombinant MKK6 can be produced in
CC transformed host cells. MKKs (W06318-22) are useful in the
CC treatment of MKK-related disorders, e.g. ischemic heart disease
CC and kidney failure, for identifying modulators of MKK activity,
CC and for raising antibodies.
SQ Sequence 334 AA;

Query Match 34.8%; Score 48; DB 1; Length 334;
Best Local Similarity 37.5%; Pred. No. 8.2;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLIAKNNKAS 28
|:::||||: ||| | : |
DB 151 PEDILGKIAVSIVKALEHLRSKLS 174

RESULT 7

W19631
ID W19631 standard; Protein; 334 AA.
AC W19631;
DT 01-SEP-1997 (first entry)
DE Human mitogen-activated protein kinase kinase MEK6
KW MEK-6; mitogen-activated protein kinase kinase; MAPKK; p38;
KW signal transduction; cell proliferation; osteoarthritis; ischaemia;
KW reperfusion injury; trauma; cancer; autoimmune disease;
KW rheumatoid arthritis; multiple sclerosis; psoriasis;
KW inflammatory bowel disease; therapy; diagnosis.
OS Homo sapiens.
PN W09722704-A1.
PD 26-JUN-1997.
PF 20-DEC-1996; U20233.
PR 20-DEC-1995; US-576240.
PA (SIGN-) SIGNAL PHARM INC.
PI Stein B, Yang MXH;
DR WPI: 97-341695/31.
DR N-PSDB; T68716.
PT Mitogen-activated protein kinase kinase and corresponding DNA -
PT phosphorylates p38 cascade members, used in the treatment of
PT immunological-related cell proliferative diseases
PS Claim 1; Page 31-33; 67pp; English.
CC Human mitogen-activated protein kinase kinase MEK6 (W19631) is a
CC protein capable of modulating the activity of the mitogen-activated
CC protein kinase p38 (esp. p38-2). Its amino acid sequence was
CC deduced from a cDNA clone (T68716) obtd. from a MOLT-4 cDNA
CC library. MEK6 is 88% identical to its closest homologue MKK3,
CC and all relevant kinase subdomains are conserved. MEK6
CC polypeptides can be produced in transformed or transfected host
CC cells. MEK6 and its variants, antibodies raised against MEK6, and
CC MEK6 nucleic acids can be used to modulate (stimulate or inhibit)
CC phosphorylation of p38 by MEK6 for use in the treatment of diseases
CC associated with the p38 cascade e.g. immunological-related cell
CC proliferative diseases and autoimmune diseases, and also to detect
CC MEK6 kinase activity and identify proteins that interact with MEK6.
SQ Sequence 334 AA;

Query Match 34.8%; Score 48; DB 1; Length 334;
Best Local Similarity 37.5%; Pred. No. 8.2;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLIAKNNKAS 28
|:::||||: ||| | : |
DB 151 PEDILGKIAVSIVKALEHLRSKLS 174

RESULT 8

W97664
ID W97664 standard; Protein; 260 AA.
AC W97664;
DT 10-MAY-1999 (first entry)
DE Human mitogen activated protein kinase kinase MKK7.
KW MKK7; mitogen activated protein kinase kinase; MAP kinase kinase;
KW human; signal transduction; inflammation; ischaemic heart disease;
KW kidney failure; oxidative liver damage; burn; septic shock;
KW respiratory distress syndrome; rheumatoid arthritis;
KW autoimmune disorder; therapy.
OS Homo sapiens.
PN W09902547-A1.
PD 21-JAN-1999.

DT 10-MAY-1999 (first entry)
DE Human mitogen activated protein kinase kinase MKK6.
KW MKK6; mitogen activated protein kinase kinase; MAP kinase
KW human; signal transduction; inflammation; psoriasis; AIDS; cancer;
KW apoptosis; therapy.
OS Homo sapiens.
PN W09902547-A1.
PD 21-JAN-1999.
PR 07-JUL-1998; U14101.
PR 07-JUL-1997; US-888429.
PA (UYMA-) UNIV MASSACHUSETTS.
PI Davis RJ, Tournier C, Whitmarsh A;
DR WPI; 99-120771/10.
DR N-PSDB; X07066.
PT New isolated mitogen-activated protein kinase isoforms - used
PT to develop products for treating e.g. inflammatory disorders,
PT oxidative damage, proliferative disorders or autoimmune disorders.
PS Disclosure; Page 129; 168pp; English.
CC This polypeptide comprises human mitogen activated protein (MAP)
CC kinase kinase 6 (MKK6). The amino acid sequence was deduced from an
CC isolated human skeletal muscle cDNA clone (see X07066) and shows
CC high homology to human MKK3 (see W97668). The human MAP kinase
CC kinases MKK3, MKK4, MKK6 and MKK7 (see W97664 and W97668-72),
CC described in the invention, mediate the transduction of specific
CC signals from the cell surface to the nucleus along specific
CC pathways. They are useful for screening reagents which modulate
CC MKK activity. Such agents can be used to prevent or treat
CC MKK-mediated disorders, e.g. inflammation, oxidative damage or
CC stress-related proliferative disorders, e.g. psoriasis, AIDS,
CC malignancies of e.g. the skin, bone marrow, lung, liver, breast,
CC gastrointestinal system and genito-urinary tract. Agents which
CC inhibit the activity or expression of MKK inhibit cell growth or
CC cause apoptosis. MKK7 polynucleotides (see X07059-64) and
CC polypeptides (see W97662-67) are claimed.
CC Sequence 334 AA;
SQ

Query Match 34.8%; Score 48; DB 1; Length 334;
Best Local Similarity 37.5%; Pred. No. 8.2;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps

QY 5 PDKVLGKVAEVLRLVIKKNKAS 28
| : ||||| :: | : | |
Db 151 PEDILGKIANSVKALEHLSKLS 174

RESULT 13
W97663
ID W97663 standard; Protein; 389 AA.
AC AC W97663;
DT 10-MAY-1999 (first entry)
DE Mouse mitogen activated protein kinase kinase MKK7b splice variant.
KW MKK7b; mitogen activated protein kinase kinase; MAP kinase kinase;
KW mouse; signal transduction; inflammation; ischaemic heart disease;
KW kidney failure; oxidative liver damage; burn; septic shock;
KW respiratory distress syndrome; rheumatoid arthritis;
KW autoimmune disorder; therapy; splice variant.
OS Mus sp.
PN W09902547-A1.
PD 21-JAN-1999.
PR 07-JUL-1998; U14101.
PR 07-JUL-1997; US-888429.
PA (UYMA-) UNIV MASSACHUSETTS.
PI Davis RJ, Tournier C, Whitmarsh A;
DR WPI; 99-120771/10.
DR N-PSDB; X07060.
PT New isolated mitogen-activated protein kinase isoforms - used
PT to develop products for treating e.g. inflammatory disorders,
PT oxidative damage, proliferative disorders or autoimmune disorders
PS Claim 2; Page 148; 168pp; English.
CC This polypeptide comprises murine mitogen activated protein (MAP)
CC kinase kinase 7 splice variant MKK7b (missing the N-terminal end),
CC a protein that has serine, threonine and tyrosine kinase activity,

Query Match 34.8%; Score 48; DB 1; Length 380;
 Best Local Similarity 36.4%; Pred. No. 9.4;
 Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0

QY 5 PDKVLGKVAEVLRLVLIAKKNK 26
 |:::|::|:| |:::| | | | |
 Db 142 PERILGKMTVAIVKALYYLKEK 163

RESULT 11
 W97667

ID W97667 standard; Protein; 453 AA.
 AC W97667;
 DE 10-MAY-1999 (first entry)
 DE Mouse mitogen activated protein kinase MKK7e splice variant.
 KW MKK7e; Mitogen activated protein kinase kinase; MAP kinase kinase;
 KW mouse; signal transduction; inflammation; ischaemic heart disease;
 KW kidney failure; oxidative liver damage; burn; septic shock;
 KW respiratory distress syndrome; rheumatoid arthritis;
 KW autoimmune disorder; therapy; splice variant.
 MS Mus sp.
 PN W09502547-A1.
 PS 21-JAN-1999.
 PD 07-JUL-1998; U14101.
 PF 07-JUL-1997; US-888429.
 PP (UYMA-) UNIV MASSACHUSETTS.
 PA Davis RJ, Tournier C, Whitmarsh A;
 PI WPI: 99-120771/10.
 DR N-PSDB: X07064.
 DR New isolated mitogen-activated protein kinase isoforms - used
 PPT to develop products for treating e.g. inflammatory disorders,
 PPT oxidative damage, proliferative disorders or autoimmune disorders
 PPT Claim 2: Page 161-162; 168pp; English.
 PS This polypeptide comprises the full-length murine mitogen activated
 CC protein (MAP) kinase kinase 7 splice variant MKK7e, a protein that
 CC has serine, threonine and tyrosine kinase activity, and which
 CC phosphorylates JNK protein kinase, but not p38. The amino acid
 CC sequence of MKK7d was deduced from a cDNA clone (see X07064)
 CC isolated from a mouse testis cDNA library. Mouse and human MKK7
 CC isoforms (see W97662-67) have been identified. These are useful
 CC for screening reagents which modulate MKK activity. Such agents
 CC can be used to prevent or treat MKK-mediated disorders, e.g.
 CC inflammation, oxidative damage or stress-related proliferative
 CC disorders, e.g. psoriasis, AIDS, malignancies of e.g. the skin,
 CC bone marrow, lung, liver; breast, gastrointestinal system and
 CC genito-urinary tract. Agents which inhibit the activity or
 CC expression of MKK inhibit cell growth or cause apoptosis. MKK7
 CC and agents that modulate its activity can be used to treat ischemic
 CC heart disease, kidney failure, oxidative liver damage, respiratory
 CC distress syndrome, heat and radiation burns, septic shock,
 CC rheumatoid arthritis, autoimmune disorders, or inflammatory
 CC diseases (claimed). The products can also be used for detection,
 CC diagnosis and production of transgenic animals. MKK7
 CC polynucleotides can be used in claimed hybridization assays to
 CC detect RNA encoding MKK7 in a cell.
 CC Sequence 453 AA;
 SQ

Query Match 34.8%; Score 48; DB 1; Length 453;
 Best Local Similarity 36.4%; Pred. No. 11;
 Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0

QY 5 PDKVLGKVAEVLRLVLIAKKNK 26
 |:::|::|:| |:::| | | | |
 Db 215 PERILGKMTVAIVKALYYLKEK 236

RESULT 12
 W97669

ID W97669 standard; Protein; 334 AA.
 AC W97669;

CC and which phosphorylates JNK protein kinase, but not p38. The
CC amino acid sequence was deduced from a cDNA clone (see X07060)
CC isolated from a mouse testis cDNA library. Mouse and human MKK7
CC isoforms (see W97662-67) have been identified. These are useful
CC for screening reagents which modulate MKK activity. Such agents
CC can be used to prevent or treat MKK-mediated disorders, e.g.
CC inflammation, oxidative damage or stress-related proliferative
CC disorders, e.g. psoriasis, AIDS, malignancies of e.g. the skin,
CC bone marrow, lung, liver, breast, gastrointestinal system and
CC genito-urinary tract. Agents which inhibit the activity or
CC expression of MKK inhibit cell growth or cause apoptosis. MKK7
CC and agents that modulate its activity can be used to treat ischemic
CC heart disease, kidney failure, oxidative liver damage, respiratory
CC distress syndrome, heat and radiation burns, septic shock,
CC rheumatoid arthritis, autoimmune disorders, or inflammatory
CC diseases (claimed). The products can also be used for detection,
CC diagnosis and production of transgenic animals. MKK7
CC polynucleotides can be used in claimed hybridization assays to
CC detect RNA encoding MKK7 in a cell.
CC Sequence 389 AA;
SQ

Query Match 34.8%; Score 48; DB 1; Length 389;
Best Local Similarity 36.4%; Pred. No. 9.6;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 PDKVLGKVAEVLRLIAKKNK 26
|:::||||:|:::| | |
Db 185 PERILGKMTVAIVKALYLYLKEK 206

RESULT 14
ID W94853 standard; Protein; 468 AA.
AC W94853;
DT 10-MAY-1999 (first entry)
DE Mouse mutant MAPK kinase MKK7.
KW Mitogen-activated protein kinase; MAPK kinase; MAPK; MKK7; SAPK/JNK;
KW stress activated protein kinase; c-Jun N-terminal kinase; SAPK; JNK;
KW Fas antigen; graft-versus-host disease; toxic epidermal necrolysis;
KW lupus; IgA kidney disease; gene therapy; p38; TNF-alpha; mouse; mutant.
OS Mus sp.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc_difference 165
FT Key /label= K165L
FT /note= "wild-type Lys is replaced by Leu"
PN W09901559-AL.
PD 14-JAN-1999.
PF 03-JUL-1998; J03016.
PR 03-JUL-1997; JP-193207.
PA (ASAH) ASAH KASEI KOGYO KK.
PI Matsuzaki O, Moriguchi T, Nishida E;
DR WPI: 99-106059/09.
DR N-PSDB: X05982.
PT New mitogen activated protein kinase of vertebrate origin -
PT activates SAPK/JNK (but not p38) stimulation in response to Fas
PT antigen or TNF-alpha, used in, e.g. gene therapy
PS Claim 8; Page 80-83; 92pp; Japanese.
CC The invention relates to a novel mitogen-activated protein kinase (MAPK)
CC kinase, designated MKK7 of vertebrate origin and widely expressed in
CC tissues. The invention provides nucleic acid sequences encoding human
CC stress activated protein kinase; c-Jun N-terminal kinase; SAPK; JNK;
CC Fas antigen; graft-versus-host disease; toxic epidermal necrolysis;
CC lupus; IgA kidney disease; gene therapy; p38; TNF-alpha; mouse; mutant.
CC Mus sp.
CC Synthetic.
CC Key
CC Misc_difference 165
CC Key /label= K165L
CC /note= "wild-type Lys is replaced by Leu"
PN W09901559-AL.
PD 14-JAN-1999.
PF 03-JUL-1998; J03016.
PR 03-JUL-1997; JP-193207.
PA (ASAH) ASAH KASEI KOGYO KK.
PI Matsuzaki O, Moriguchi T, Nishida E;
DR WPI: 99-106059/09.
DR N-PSDB: X05982.
PT New mitogen activated protein kinase of vertebrate origin -
PT activates SAPK/JNK (but not p38) stimulation in response to Fas
PT antigen or TNF-alpha, used in, e.g. gene therapy
PS Claim 8; Page 80-83; 92pp; Japanese.
CC The invention relates to a novel mitogen-activated protein kinase (MAPK)
CC kinase, designated MKK7 of vertebrate origin and widely expressed in
CC tissues. The invention provides nucleic acid sequences encoding human
CC stress activated protein kinase; c-Jun N-terminal kinase; SAPK; JNK;
CC Fas antigen; graft-versus-host disease; toxic epidermal necrolysis;
CC lupus; IgA kidney disease; gene therapy; p38; TNF-alpha; mouse; mutant.
CC Mus sp.
CC Synthetic.
CC Key
CC Misc_difference 165
CC Key /label= K165L
CC /note= "wild-type Lys is replaced by Leu"

CC a mouse mutant MKK7.
SQ Sequence 468 AA;

Query Match 34.8%; Score 48; DB 1; Length 468;
Best Local Similarity 36.4%; Pred. No. 12;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 PDKVLGKVAEVLRLIAKKNK 26
|:::||||:|:::| | |
Db 231 PERILGKMTVAIVKALYLYLKEK 252

RESULT 15
ID W94849 standard; Protein; 435 AA.
AC W94849;
DT 12-MAY-1999 (first entry)
DE Human MAPK kinase MKK7.
KW Mitogen-activated protein kinase; MAPK kinase; MAPK; MKK7; SAPK/JNK;
KW stress activated protein kinase; c-Jun N-terminal kinase; SAPK; JNK;
KW Fas antigen; graft-versus-host disease; toxic epidermal necrolysis;
KW lupus; IgA kidney disease; gene therapy; p38; TNF-alpha; human.
OS Homo sapiens.
PN W09901559-AL.
PD 14-JAN-1999.
PF 03-JUL-1998; J03016.
PR 03-JUL-1997; JP-193207.
PA (ASAH) ASAH KASEI KOGYO KK.
PI Matsuzaki O, Moriguchi T, Nishida E;
DR WPI: 99-106059/09.
DR N-PSDB: X05978.
PT New mitogen activated protein kinase of vertebrate origin -
PT activates SAPK/JNK (but not p38) stimulation in response to Fas
PT antigen or TNF-alpha, used in, e.g. gene therapy
PS Claim 2; Page 59-62; 92pp; Japanese.
CC The invention relates to a novel mitogen-activated protein kinase (MAPK)
CC kinase, designated MKK7 of vertebrate origin and widely expressed in
CC tissues. The invention provides nucleic acid sequences encoding human
CC stress activated protein kinase; c-Jun N-terminal kinase; SAPK; JNK;
CC Fas antigen; graft-versus-host disease; toxic epidermal necrolysis;
CC lupus; IgA kidney disease; gene therapy; p38; TNF-alpha; mouse; mutant.
CC Mus sp.
CC Synthetic.
CC Key
CC Misc_difference 165
CC Key /label= K165L
CC /note= "wild-type Lys is replaced by Leu"

Query Match 34.8%; Score 48; DB 1; Length 435;
Best Local Similarity 36.4%; Pred. No. 11;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 PDKVLGKVAEVLRLIAKKNK 26
|:::||||:|:::| | |
Db 231 PERILGKMTVAIVKALYLYLKEK 252

Search completed: November 13, 1999, 18:59:46
Job time: 125 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	51	37.0	400	1	US-08-530-950-12	Sequence 12, Appl	
2	51	37.0	400	2	US-08-878-983-3	Sequence 3, Appl	
3	49	35.5	400	2	US-08-878-983-17	Sequence 17, Appl	
4	48	34.8	334	1	US-08-530-950-4	Sequence 4, Appl	
5	47	34.1	318	1	US-08-530-950-2	Sequence 2, Appl	
6	47	34.1	318	2	US-08-446-083-2	Sequence 2, Appl	
7	46	33.3	393	1	US-08-423-3989-33	Sequence 33, Appl	
8	46	33.3	393	1	US-08-530-950-11	Sequence 11, Appl	
9	43.5	31.5	486	2	US-08-870-518-3	Sequence 3, Appl	
10	42	30.4	583	2	US-08-448-196A-5	Sequence 5, Appl	
11	42	30.4	344	2	US-08-726-575A-2	Sequence 2, Appl	
12	41	29.7	604	2	US-08-756-317-6	Sequence 6, Appl	
13	41	29.7	2154	2	US-08-841-349-4	Sequence 4, Appl	
14	40.5	29.3	2339	2	US-08-455-543A-47	Sequence 47, Appl	
15	40.5	29.3	2327	2	US-08-455-543A-48	Sequence 48, Appl	
16	40.5	29.3	2339	2	US-08-223-305C-47	Sequence 47, Appl	
17	40.5	29.3	2327	2	US-08-223-305C-48	Sequence 48, Appl	
18	40	29.0	393	1	US-08-530-950-8	Sequence 8, Appl	
19	40	29.0	1101	2	US-08-916-917-14	Sequence 14, Appl	
20	40	29.0	1049	2	US-08-817-090B-2	Sequence 2, Appl	
21	40	29.0	1050	2	US-08-817-090B-4	Sequence 4, Appl	
22	39.5	28.6	648	1	US-08-276-151-2	Sequence 2, Appl	
23	39.5	28.6	346	1	US-08-276-151-5	Sequence 5, Appl	
24	39.5	28.6	648	1	US-08-185-283-1	Sequence 1, Appl	
25	39.5	28.6	648	1	US-08-185-283-2	Sequence 2, Appl	
26	39.5	28.6	648	1	US-08-185-283-3	Sequence 3, Appl	
27	39.5	28.6	648	1	US-08-185-283-4	Sequence 4, Appl	
28	39.5	28.6	648	1	US-08-185-283-5	Sequence 5, Appl	
29	39.5	28.6	648	1	US-08-185-283-12	Sequence 12, Appl	
30	39.5	28.6	2237	1	US-08-354-973-1	Sequence 1, Appl	
31	39.5	28.6	648	2	US-08-867-751A-6	Sequence 6, Appl	
32	39.5	28.6	267	2	US-07-857-224B-42	Sequence 42, Appl	
33	39	28.3	276	1	US-07-612-674-5	Sequence 5, Appl	
34	39	28.3	1023	1	US-08-198-445B-4	Sequence 4, Appl	
35	39	28.3	2368	1	US-08-198-446B-15	Sequence 15, Appl	
36	39	28.3	307	1	US-08-713-828-1	Sequence 1, Appl	
37	39	28.3	307	2	US-08-919-627-1	Sequence 1, Appl	
38	39	28.3	1023	2	US-08-870-693-4	Sequence 4, Appl	
39	39	28.3	2368	2	US-08-870-693-15	Sequence 15, Appl	

APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl G.
APPLICANT: Lal, Preeti
APPLICANT: Goli, Surya K.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,989
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0321 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TMLR3DT01
CLONE: 402339
US-08-878-989-3

Query Match 37.0%; Score 51; DB 2; Length 400;
Best Local Similarity 45.5%; Pred. No. 1;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 5 PDKVLGKVAVEVLRLIAKNK 26
|:|||||: ||| | : |
Db 166 PEELGKVSIAVLRLGLAYLREK 187

RESULT 3
US-08-878-989-17
Sequence 17, Application US/08878989
Patent No. 5885803
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl G.
APPLICANT: Lal, Preeti
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,989
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0321 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 854170
US-08-878-989-17

Query Match 35.5%; Score 49; DB 2; Length 400;
Best Local Similarity 45.5%; Pred. No. 2.2;
Matches 10; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 5 PDKVLGKVAVEVLRLIAKNK 26
|:|||||: ||| | : |
Db 166 PEELGKVSIAVLRLGLAYLREK 187

RESULT 4
US-08-530-950-4
Sequence 4, Application US/08530950
Patent No. 5736381
GENERAL INFORMATION:
APPLICANT: Davis, Roger J.
APPLICANT: Raingeaud, Joel
APPLICANT: Gupta, Shashi
APPLICANT: Derijard, Benoit
TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
TITLE OF INVENTION: KINASES
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,950

FILING DATE: 19-SEP-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Fasse, J. Peter
 REGISTRATION NUMBER: 32,983
 REFERENCE/DOCKET NUMBER: 07917/010001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/543-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 334 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 US-08-530-950-4

Query Match 34.8%; Score 48; DB 1; Length 334;
Best Local Similarity 37.5%; Pred. No. 2.6;
Matches 9; Conservative 7; Mismatches 8; Indels

QY 5 PDKVLGKVAEVLRLVLI AKNKAS 28
| : ||| : || : : | : | |
Db 151 PEDILGKIAVSIVKALEHLHLSKLS 174

```

RESULT      5
US-08-530-950-2
; Sequence 2, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derifard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
;

```

Query Match 34.1%; Score 47; DB 1; Length 318;
Best Local Similarity 37.5%; Pred. No. 3.5;
Matches 9; Conservative 7; Mismatches 8; Indels

QY 5 PDKVLGKVAVEVLRVLIAKNKAS 28
| : : | : : | : : | : |
Db 133 PEDILGEIAVSIVRALEHLHLSKLS 156

RESULT 6
 US-08-446-083-2
 ; Sequence 2, Application US/08446083
 ; Patent No. 5804427
 ; GENERAL INFORMATION:
 ; APPLICANT: Davis, Roger J.
 ; APPLICANT: Raingeaud, Joel
 ; APPLICANT: Gupta, Shashi
 ; APPLICANT: Derlijard, Benoit
 ; TITLE OF INVENTION: CYTOKINE-,STRESS- AND
 ; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
 ; TITLE OF INVENTION: KINASES
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; SOURCE APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/446,083
 ; FILING DATE: 19-MAY-1995
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fasse, J. Peter
 ; REGISTRATION NUMBER: 32,983
 ; REFERENCE/DOCKET NUMBER: 04020/066001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/542-5070
 ; TELEFAX: 617/542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 318 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: both
 ; MOLECULE TYPE: protein
 US-08-446-083-2

Query Match 34.1%; Score 47; DB 2; Length 318;
Best Local Similarity 37.5%; Pred. NO. 3.5;
Matches 9; Conservative 7; Mismatches 8; Indels

```
Qy 5 PDKVLGKVAEVLRLVIAKKNKAS 28
    | : : : : : : : : : : : :
Db 133 PEDILGEIAVSIVRALEHLHLSKLS 156
```

```

RESULT      7
US-08-423-399B-33
; Sequence 33, Application US/08423399B
; Patent No. 5663314
; GENERAL INFORMATION:
; APPLICANT: Seger, R., Seger, D., Ahn, N.G., and Krebs, E.G.
; TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION MAPK KINASE
;

```


NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage
COMPUTER: IBM PC/386 Compatible
OPERATING SYSTEM: MS-DOS 4.01
SOFTWARE: Word for Windows-t
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/423,399B
FILING DATE: April 18, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/980,608
FILING DATE: NO. 5663314ember 20, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: WRFO-1-8427
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0712 (direct)
TELEFAX: 1-206-224-0779
TELEX: 4938023
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 393 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: polypeptide
DESCRIPTION: MAPK1a protein; FIGURE 10
US-08-423-399B-33

Query Match 33.3%; Score 46; DB 1; Length 393;
Best Local Similarity 36.4%; Pred. No. 6.5;
Matches 8; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLIAKKNK 26
|:::||||:|::|::|
DB 162 PEQILGKVSIAVIKGLTYLREK 183

RESULT 8
US-08-530-950-11
Sequence 11, Application US/08530950
Patent No. 5736381
GENERAL INFORMATION:
APPLICANT: Davis, Roger J.
APPLICANT: Raugeaud, Joel
APPLICANT: Gupta, Shashi
APPLICANT: Derjard, Benoit
TITLE OF INVENTION: CYCOKINE-, STRESS-, AND
TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30,
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/530,950
FILING DATE: 19-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07917/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 393 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
US-08-530-950-11

Query Match 33.3%; Score 46; DB 1; Length 393;
Best Local Similarity 36.4%; Pred. No. 6.5;
Matches 8; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLIAKKNK 26
|:::||||:|::|::|
DB 162 PEQILGKVSIAVIKGLTYLREK 183

RESULT 9
US-08-870-518-3
Sequence 3, Application US/08870518
Patent No. 5925566
GENERAL INFORMATION:
APPLICANT: Davis, Roger J.
APPLICANT: Galcheva-Gargova, Zoya
TITLE OF INVENTION: NON-ACTIVATED RECEPTOR COMPLEX
TITLE OF INVENTION: PROTEINS AND USES THEREOF
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870,518
FILING DATE: 06-JUN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,219
FILING DATE: 06-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, Peter J.
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/102001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-870-518-3

```
Query Match 31.5%; Score 43.5; DB 2; Length 486;
Best Local Similarity 40.7%; Pred. No. 21;
Matches 11; Conservative 7; Mismatches 6; Indels 3; Gaps 1;

QY 2 IYKPKVLGKRVAVELRLVLIANKKAS 28
Db 8 LFRP---VGEAAAEVDESTAEQNKAN 31

RESULT 10
US-08-448-196A-5
; Sequence 5, Application US/08448196A
; Patent No. 5780594
; GENERAL INFORMATION:
; APPLICANT: CARTER, DANIEL C.
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE PROTEIN FRAGMENTS
; TITLE OF INVENTION: CONTAINING SPECIFIC BINDING REGIONS OF SERUM ALBUMIN OR
; TITLE OF INVENTION: RELATED PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NASA
; STREET: MARSHALL SPACE FLIGHT CENTER
; CITY: HUNTSVILLE
; STATE: ALABAMA
; COUNTRY: USA
; ZIP: 35812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,196A
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: BROAD JR., ROBERT L.
; REGISTRATION NUMBER: 18,757
; REFERENCE/DOCKET NUMBER: XX/MFS-28402-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 205-544-0021
; TELEFAX: 205-544-0258
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-448-196A-5

Query Match 30.4%; Score 42; DB 2; Length 583;
Best Local Similarity 64.3%; Pred. No. 44;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 PDKVLGKRVAVELRL 18
Db 130 PDKFLGKLYEVAR 143

RESULT 11
US-08-726-575A-2
; Sequence 2, Application US/08726575A
; Patent No. 5834587
; GENERAL INFORMATION:
; APPLICANT: Winnie Chan, Derk J. Bergsma,
; APPLICANT: Catherine E. Ellis
; TITLE OF INVENTION: A No. 5834587el G-Protein Coupled Receptor,
; TITLE OF INVENTION: HLTX11
```

```
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road, P.O. Box 1539
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726.575A
; FILING DATE: OCTOBER 8, 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG 50025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610 270 5219
; TELEFAX: 610 270 4026
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-726-575A-2

Query Match 30.4%; Score 42; DB 2; Length 344;
Best Local Similarity 42.3%; Pred. No. 25;
Matches 11; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 2 IYKPKVLGKRVAVELRLVLIANKKKA 27
Db 49 IFIPGLLSAALWVLCRFISKKKA 74

RESULT 12
US-08-756-317-6
; Sequence 6, Application US/08756317
; Patent No. 5849894
; GENERAL INFORMATION:
; APPLICANT: Clemente, Thomas E.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Mitsky, Timothy A.
; APPLICANT: Stark, David M.
; TITLE OF INVENTION: Improved Rhodospirillum Rubrum
; TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,317
; FILING DATE: 25-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
```

RESULT 15
US-08-455-343A-48
Sequence 48, Application US/08455343A
Patent No. 5792846
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: Feldman, Daniel
APPLICANT: McCue, Ann
APPLICANT: Brenner, Robert

;; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND

;; TITLE OF INVENTION: METHODS

;; NUMBER OF SEQUENCES: 57

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Brown, Martin, Haller & McClain

;; STREET: 1660 Union Street

;; CITY: San Diego

;; STATE: California

;; COUNTRY: USA

;; ZIP: 92101-2926

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette

;; COMPUTER: IBM Compatible

;; OPERATING SYSTEM: DOS

;; SOFTWARE: FastSeq Version 1.5

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/455,543A

;; FILING DATE: May 31, 1995

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/223,305

;; FILING DATE: April 4, 1994

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 07/868,354

;; FILING DATE: April 10, 1992

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/745,206

;; FILING DATE: 15-AUG-1991

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/620,250

;; FILING DATE: 30-NOV-1990

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/482,384

;; FILING DATE: 20-FEB-1990

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/603,751

;; FILING DATE: 04-APR-1989

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: WO PCT/US89/01408

;; FILING DATE: 04-APR-1989

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/176,899

;; FILING DATE: 04-APR-1988

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Seidman, Stephanie L.

;; REGISTRATION NUMBER: 33,779

;; REFERENCE/DOCKET NUMBER: 6362-52517

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (619)238-0999

;; TELEFAX: (619)238-0062

;; INFORMATION FOR SEQ ID NO: 48:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 2237 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; FRAGMENT TYPE: internal

;; US-08-455-543A-48

Query Match 29.3%; Score 40.5; DB 2; Length 2237;

Best Local Similarity 35.7%; Pred. No. 3.4e+02;

Matches 10; Conservative 8; Mismatches 9; Indels 1; Gaps 1;

QY 3 YKPKV-LGKVAVEVLRLIAKNKASR 29

Db 1844 HKPDMTVGKVYAALMIFDYKQNKTR 1871

Search completed: November 13, 1999, 10:56:27

Job time: 1356 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:32 ; Search time 251.81 Seconds
(without alignments)
7.291 Million cell updates/sec

Title: US-08-913-430-4

Perfect score: 138

Sequence: 1 TTYKKDKVLGVAVFVRLVLIKKKASR 29

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA:*
1: /cgn2.6/ptodata/1/paa/PTUS9_COMB.pep.*
2: /cgn2.6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2.6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2.6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2.6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2.6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2.6/ptodata/1/paa/US084A_COMB.pep.*
8: /cgn2.6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2.6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2.6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2.6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2.6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2.6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2.6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2.6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2.6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2.6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2.6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2.6/ptodata/1/paa/US096_COMB.pep.*
20: /cgn2.6/ptodata/1/paa/PT_NEW_COMB.pep.*
21: /cgn2.6/ptodata/1/paa/US07_NEW_COMB.pep.*
22: /cgn2.6/ptodata/1/paa/US08_NEW_COMB.pep.*
23: /cgn2.6/ptodata/1/paa/US09_NEW_COMB.pep.*
24: /cgn2.6/ptodata/1/paa/US09_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	138	100.0	419	4	US-08-183-774-1
2	138	100.0	419	13	US-08-913-430-2
3	138	100.0	29	13	US-08-913-430-4
4	51	37.0	400	1	PCT-US96-01078-12
5	51	37.0	400	12	US-08-888-429-12
6	51	37.0	400	12	US-08-888-429-12
7	51	37.0	259	12	US-08-888-518-31
8	48	34.8	334	1	PCT-US96-01078-4
9	48	34.8	334	9	US-08-576-240-2
10	48	34.8	334	12	US-08-888-429-4
11	48	34.8	389	12	US-08-888-429-20
12	48	34.8	260	12	US-08-888-429-26
13	48	34.8	419	12	US-08-888-429-28
14	48	34.8	380	12	US-08-888-429-30
15	48	34.8	453	12	US-08-888-429-32
16	48	34.8	334	12	US-08-888-429-34
17	48	34.8	389	12	US-08-888-429A-4
18	48	34.8	260	12	US-08-888-429A-20
19	48	34.8	419	12	US-08-888-429A-26
					Sequence 28, Appl

20	48	34.8	380	12	US-08-888-429A-30
21	48	34.8	453	12	US-08-888-429A-32
22	47	34.1	318	1	PCT-US96-01078-2
23	47	34.1	183	12	US-08-827-356-4667
24	47	34.1	318	12	US-08-888-429-2
25	47	34.1	318	12	US-08-888-429A-2
26	47	34.1	318	12	US-08-888-429A-2
27	47	34.1	318	16	US-09-233-087-2
28	46	33.3	393	1	PCT-US96-01078-11
29	46	33.3	393	1	PCT-US97-19575-79
30	46	33.3	393	2	US-07-980-608A-33
31	46	33.3	393	2	US-07-980-608B-33
32	46	33.3	430	11	US-08-759-739-358
33	46	33.3	429	11	US-08-759-739-456
34	46	33.3	393	12	US-08-888-429-11
35	46	33.3	393	12	US-08-888-429A-11
36	46	33.3	480	13	US-08-987-367-4
37	46	33.3	762	16	US-09-228-986-114
38	45	32.6	393	12	US-08-888-429-21
39	45	32.6	393	12	US-08-888-429A-21
40	45	32.6	281	15	US-09-134-000-3822
41	44.5	32.2	324	16	US-09-134-000-3822
42	44.5	32.2	324	19	US-09-248-796-20010
43	44	31.9	127	1	PCT-US97-02318-197
44	44	31.9	146	8	US-08-487-032A-500
45	44	31.9	146	9	US-08-561-469A-500

ALIGNMENTS

RESULT 1
US-08-183-774-1
; Sequence 1. Application US/08183774
; GENERAL INFORMATION:
; APPLICANT: SEIO, Yasuhiro
; APPLICANT: FUTU, Satoshi
; APPLICANT: MITSUSE, Shizuo
; APPLICANT: MATSUO, Kanako
; APPLICANT: TSUNA, Mika
; TITLE OF INVENTION: DNA's Encoding Surface Antigen of
; TITLE OF INVENTION: Mycoplasma hyopneumoniae, DNA Fragments for Primer,
; TITLE OF INVENTION: Recombinant Antigenic Peptides and Diagnostic Method of
; TITLE OF INVENTION: Mycoplasma pneumoniae of Swine Using Same
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/183,774
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/747,015
; FILING DATE: 19-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, Norman F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4183-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:


```

; APPLICANT: Whitmarsh, Alan
; APPLICANT: Tournier, Cathy
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN- ACTIVATED HUMAN PROTEIN KINASE KINASES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,429
; FILING DATE: 07-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,950
; FILING DATE: 19-SEP-1995
; APPLICATION NUMBER: 08/446,083
; FILING DATE: 19-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 299354
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-888-429-12

Query Match 37.0%; Score 51; DB 12; Length 400;
Best Local Similarity 45.5%; Pred. No. 9.1;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLVLIKKNK 26
|:::||||: ||| | : |
Db 166 PEEILGKVSIAVLRLGLAYLREK 187

RESULT 6
US-08-888-429A-12
; Sequence 12, Application US/08888429A
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Whitmarsh, Alan
; APPLICANT: Tournier, Cathy
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-
; TITLE OF INVENTION: ACTIVATED HUMAN PROTEIN KINASE KINASES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,429A
; FILING DATE: 07-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,950
; FILING DATE: 19-SEP-1995
; APPLICATION NUMBER: 08/446,083
; FILING DATE: 19-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 299354
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-888-429A-12

Query Match 37.0%; Score 51; DB 12; Length 400;
Best Local Similarity 45.5%; Pred. No. 9.1;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLVLIKKNK 26
|:::||||: ||| | : |
Db 166 PEEILGKVSIAVLRLGLAYLREK 187

RESULT 7
US-08-889-518-31
; Sequence 31, Application US/08889518B
; GENERAL INFORMATION:
; APPLICANT: Pleiman, Christopher M
; TITLE OF INVENTION: NOVEL KDS MOLECULES AND USES RELATED THERETO
; FILE REFERENCE: CPI-011
; CURRENT APPLICATION NUMBER: US/08/889,518B
; CURRENT FILING DATE: 1997-07-08
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 31
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-889-518-31

Query Match 37.0%; Score 51; DB 12; Length 259;
Best Local Similarity 45.5%; Pred. No. 5.4;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLVLIKKNK 26
|:::||||: ||| | : |
Db 108 PEEILGKVSIAVLRLGLAYLREK 129

RESULT 8
PCT-US96-01078-4
; Sequence 4, Application PCT/TUS9601078
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16

```



```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/01078
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; PCT-US96-01078-4

Query Match 34.8%; Score 48; DB 1; Length 334;
Best Local Similarity 37.5%; Pred. No. 22;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLVLIANKKAS 28
   | : ||| : || : | : |
Db 151 PEDILGKIAVSIVKALEHLHLSKLS 174

RESULT 9
US-08-576-240-2
; Sequence 2, Application US/08576240
; GENERAL INFORMATION:
; APPLICANT: Stein, Bernd
; APPLICANT: Yang, Maria
; TITLE OF INVENTION: MITOGEN-ACTIVATED PROTEIN KINASE KINASE
; TITLE OF INVENTION: MEK6 AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/576,240
; FILING DATE: 20-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
```

```
;
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.403
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-576-240-2

Query Match 34.8%; Score 48; DB 9; Length 334;
Best Local Similarity 37.5%; Pred. No. 22;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLVLIANKKAS 28
   | : ||| : || : | : |
Db 151 PEDILGKIAVSIVKALEHLHLSKLS 174

RESULT 10
US-08-888-429-4
; Sequence 4, Application US/08888429
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Whitmarsh, Alan
; APPLICANT: Tournier, Cathy
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE KINASES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,429
; FILING DATE: 07-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,950
; FILING DATE: 19-SEP-1995
; APPLICATION NUMBER: 08/446,083
; FILING DATE: 19-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 299354
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-888-429-4

Query Match 34.8%; Score 48; DB 12; Length 334;
Best Local Similarity 37.5%; Pred. No. 22;
```

```

1  NUMBER OF SEQUENCES: 34
2  CORRESPONDENCE ADDRESS:
3  ADDRESSEE: Fish & Richardson P.C.
4  STREET: 225 Franklin Street
5  CITY: Boston
6  STATE: MA
7  COUNTRY: USA
8  ZIP: 02110-2804
9  COMPUTER READABLE FORM:
10 MEDIUM TYPE: Diskette
11 COMPUTER: IBM Compatible
12 OPERATING SYSTEM: Windows95
13 SOFTWARE: FASTSEQ for Windows Version 2.0
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/08/888,429
16 FILING DATE: 07-JUL-1997
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: 08/530,950
19 FILING DATE: 19-SEP-1995
20 APPLICATION NUMBER: 08/446,083
21 FILING DATE: 19-MAY-1995
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Fasse, Peter J.
24 REGISTRATION NUMBER: 32,983
25 REFERENCE/DOCKET NUMBER: 07917/053001
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 617/542-5070
28 TELEFAX: 617/542-8906
29 TELEX: 299354
30 INFORMATION FOR SEQ ID NO: 26:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 260 amino acids
33 TYPE: amino acid
34 TOPOLOGY: linear
35 MOLECULE TYPE: protein
36 FRAGMENT TYPE: internal
37 US-08-888-429-26
38
39 Query Match 34.8%; Score 48; DB 12; Length 260;
40 Best Local Similarity 36.4%; Pred. No. 16;
41 Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0
42
43 QY 5 PDKVLGKVAVEVLRLVLIAKKNK 26
44 |::|||:|::| | |
45 Db 215 PERILGKTVAILKALYYIKER 236
46
47 RESULT 13
48 US-08-888-429-28
49 Sequence 28, Application US/08888429
50 GENERAL INFORMATION:
51 APPLICANT: Davis, Roger J.
52 APPLICANT: Whitmarsh, Alan
53 APPLICANT: Tournier, Cathy
54 TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
55 TITLE OF INVENTION: ONCOPROTEIN- ACTIVATED HUMAN PROTEIN KINASES
56 NUMBER OF SEQUENCES: 34
57 CORRESPONDENCE ADDRESS:
58 ADDRESSEE: Fish & Richardson P.C.
59 STREET: 225 Franklin Street
60 CITY: Boston
61 STATE: MA
62 COUNTRY: USA
63 ZIP: 02110-2804
64 COMPUTER READABLE FORM:
65 MEDIUM TYPE: Diskette
66 COMPUTER: IBM Compatible
67 OPERATING SYSTEM: Windows95
68 SOFTWARE: FASTSEQ for Windows Version 2.0
69 CURRENT APPLICATION DATA:
70 APPLICATION NUMBER: US/08/888,429
71 FILING DATE: 07-JUL-1997

```

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/530,950
;; FILING DATE: 19-SEP-1995
;; APPLICATION NUMBER: 08/446,083
;; FILING DATE: 19-MAY-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fasse, Peter J.
;; REGISTRATION NUMBER: 32,983
;; REFERENCE/DOCKET NUMBER: 07917/053001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617/542-5070
;; TELEFAX: 617/542-8906
;; TELEX: 299354
;; INFORMATION FOR SEQ ID NO: 28:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 419 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FRAGMENT TYPE: internal
;; US-08-888-429-28

Query Match 34.8%; Score 48; DB 12; Length 419;
Best Local Similarity 36.4%; Pred. No. 28;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLIAKKNK 26
|:::|:::|:::|:::|:::|
DB 215 PERILGKMTVAIVKALYYLKEK 236

RESULT 14
US-08-888-429-30
; Sequence 30, Application US/08888429
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Whitmarsh, Alan
; APPLICANT: Tournier, Cathy
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,429
; FILING DATE: 07-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,950
; FILING DATE: 19-SEP-1995
; APPLICATION NUMBER: 08/446,083
; FILING DATE: 19-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 299354
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids

;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FRAGMENT TYPE: internal
;; US-08-888-429-30

Query Match 34.8%; Score 48; DB 12; Length 380;
Best Local Similarity 36.4%; Pred. No. 25;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLIAKKNK 26
|:::|:::|:::|:::|:::|
DB 142 PERILGKMTVAIVKALYYLKEK 163

RESULT 15
US-08-888-429-32
; Sequence 32, Application US/08888429
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Whitmarsh, Alan
; APPLICANT: Tournier, Cathy
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,429
; FILING DATE: 07-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,950
; FILING DATE: 19-SEP-1995
; APPLICATION NUMBER: 08/446,083
; FILING DATE: 19-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 299354
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-888-429-32

Query Match 34.8%; Score 48; DB 12; Length 453;
Best Local Similarity 36.4%; Pred. No. 31;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLIAKKNK 26
|:::|:::|:::|:::|:::|
DB 215 PERILGKMTVAIVKALYYLKEK 236

Mon Nov 15 14:43:01 1999

us-08-913-430-4.rap

Page 7

Search completed: November 13, 1999, 05:08:33
Job time: 10859 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:07:52 ; Search time 75.45 Seconds
(without alignments)
15.400 Million cell updates/sec

Title: US-08-913-430-4
Perfect score: 138
Sequence: 1 TTYKPKVLGVAVEVLRVLIAKKNKASR 29

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	138	100.0	419	2 A56153	46K surface antige
2	56	40.6	146	2 B70142	ribosomal protein
3	55	39.9	397	1 S41054	mitogen-activated
4	52	37.7	209	2 E69641	phosphoribosyl-AMP
5	51	37.0	400	1 A46723	MAP kinase kinase
6	51	37.0	96	2 G69037	conserved hypotet
7	50	36.2	395	1 S36186	mitogen-activated
8	49.5	35.9	775	2 B64319	carbon-monoxide de
9	49	35.5	400	1 A48081	mitogen-activated
10	49	35.5	401	1 I52829	mitogen-activated
11	49	35.5	212	1 C47754	phosphoribosyl-AMP
12	48	34.8	334	2 S71631	MAP kinase kinase
13	48	34.8	334	2 S71632	MAP kinase kinase
14	48	34.8	162	2 S71633	MAP kinase kinase
15	48	34.8	332	2 C64073	MAP kinase kinase
16	48	34.8	275	2 D71087	rbp repressor homo
17	48	34.8	429	2 E64617	probable translati
18	47	34.1	337	1 A36940	hypothetical prote
19	47	34.1	318	2 A55556	sporulation-specif
20	47	34.1	177	2 A64431	MAP kinase kinase
21	46.5	33.7	405	2 A64388	hypothetical prote
22	46	33.3	393	1 A45100	3-hydroxy-3-methyl
23	46	33.3	393	1 S42068	MAP kinase kinase
24	46	33.3	393	1 JN0840	mitogen-activated
25	46	33.3	393	1 I59571	mitogen-activated
26	46	33.3	393	1 S46361	mitogen-activated
27	46	33.3	245	1 NDECR5	type II site-speci
28	46	33.3	429	2 T02512	hypothetical prote
29	46	33.3	539	2 F71896	hypothetical prote
30	45	32.6	384	2 B43592	outer membrane pro
31	44.5	32.2	1626	2 A39242	DNA topoisomerase
32	44.5	32.2	1612	2 S59969	DNA topoisomerase
33	44.5	32.2	1857	2 A48069	protein kinase MKK
34	44	31.9	187	2 S01787	fatty-acid synthas
35	44	31.9	790	2 F07025	probable relA prot
36	44	31.9	155	2 B71187	probable 50S ribos
37	44	31.9	330	2 B65179	ribose operon repr
38	44	31.9	561	2 S64068	hypothetical prote
39	44	31.9	241	2 S75347	hypothetical prote

ALIGNMENTS

RESULT 1
A56153
46K surface antigen precursor - Mycoplasma hyopneumoniae (SGC3)
C:Species: Mycoplasma hyopneumoniae
C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 09-Sep-1997
C:Accession: A56153
R:Puto, S.; Seto, Y.; Mitsuse, S.; Mori, Y.; Suzuki, T.; Kawai, K.
J. Bacteriol. 177, 1915-1917, 1995
A:Title: Molecular cloning of a 46-kilodalton surface antigen (P46) gene from Mycopla
A:Reference number: A56153; MUID:95204368
A:Accession: A56153
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <FUT>
A:Cross-references: GB:D16682; NID:g517134; PID:d1004604; PID:g1220114
C:Genetics:
A:Gene: P46
A:Genetic code: SGC3
C:Keywords: surface antigen

Query Match 100.0% Score 138; DB 2; Length 419;
Best Local Similarity 100.0% Pred. No. 7.2e-13;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTYKPKVLGVAVEVLRVLIAKKNKASR 29
|||||
Db 341 TTYKPKVLGVAVEVLRVLIAKKNKASR 369

RESULT 2
B70142
ribosomal protein L13 (rplM) - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
C:Accession: B70142
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: B70142
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-146 <KLE>
A:Cross-references: GB:AE001140; GB:AE000783; NID:g2688233; PID:g2688238; TIGR:BB0339
A:Experimental source: strain B31
C:Superfamily: Escherichia coli ribosomal protein L13

Query Match 40.6% Score 56; DB 2; Length 146;
Best Local Similarity 55.6% Pred. No. 0.31;
Matches 10; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 6 DKVLGVAVEVLRVLIAK 23
|:|:|:|:|:|:|:|:|:|
Db 28 DRILGVAVDVVKILRGK 45

RESULT 3

```

Best Local Similarity   52.4%;    Pred. NO. 1.7;
Matches 11; Conservative 3; Mismatches       7; Indels      0; Gaps     0;

QY        6 DKVLGKVAVEVLRVLIANKNK 26
          ||| | | | | | | : | : | | 
DB       146 DKLTKVGGEASEVIITAKNR 166

RESULT    5
A46723
MAP kinase kinase 2 (EC 2.7.1.1) - human
N:Alternate names: extracellular signal-regulated kinase activator kinase 2 (ERK kinase); protein kinase (EC 2.7.1.37); protein-tyrosine kinase (EC 2.7.1.112)
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 18-Nov-1994 #text_change 04-Sep-1998
C:Accession: A46723
R:Zheng, C.F.; Guan, K.L.
J. Biol. Chem. 268, 11435-11439, 1993
A:Title: Cloning and characterization of two distinct human extracellular signal-regulated kinase cDNAs
A:Reference number: A46723; MUID:93266604
A:Accession: A46723
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-400 <zhe>
A:Cross-references: GB:L11285; NID:g307184
A>Note: Sequence extracted from NCBI backbone (NCBIP:132853); this ORF is not annotated
C:Comment: This enzyme is activated by protein kinase raf-1 (see PIR:A00637). It is involved in the regulation of mitogenic activation
C:Genetics: GDB:PRKMK2; MEK2
A:Gene: GDB:PRKMK2; MEK2
A:Cross-references: GDB:202924; OMIM:601263
C:Complex: monomer
C:Function:
A:Description: catalyzes the formation of specific peptidyl-threonine-phosphate and p-pyruvate
A:Pathway: MAP kinase cascade
C:Superfamily: kinase-related transforming proteins; protein kinase homology
C:Keywords: ATP; monomer; phosphoprotein; phosphotransferase; serine/threonine/tyrosine phosphorylation
F:70-369/Domain: protein kinase homology <kln>
F:F78-86/Region: protein kinase ATP-binding motif
F:I01,I18,I94,I96/Active site: Lys, Glu, Asp, Lys #status predicted
F:F222_226/Binding site: phosphate (Ser) (covalent) (by raf-1 kinase) #status predicted
F:F394/Binding site: phosphate (Thr) (covalent) (by raf-1 kinase) #status predicted

Query Match               37.0%;    Score 51; DB 1; Length 400;
Best Local Similarity    45.5%;    Pred. No. 4.7;
Matches 10; Conservative 6; Mismatches       6; Indels      0; Gaps     0;

QY        5 PKVLGKVAVEVLRVLIANKNK 26
          ||::||:::|| | | | : | 
DB       166 PEELGKVSIIVRLGLAYLREK 187

RESULT    6
G69037
conserved hypothetical protein MTH1283 - Methanobacterium thermoautotrophicum (strain ATCC 35061)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Jul-1998
C:Accession: G69037
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwanji, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: full-genome shotgun sequencing project
A:Reference number: A69000; MUID:98037514
A:Accession: G69037
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-96 <mth>
A:Cross-references: GB:AE000893; GB:AE000666; NID:g2622375; PID:g2622390
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1283
F:8-80/Domain: hise protein similarity <HSP>
```



```

RESULT 14
S71633
MAP kinase kinase (EC 2.7.1.1) SAPKK3 isoform B - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 10-Jul-1998
C:Accession: S71633
R:Cuenda, A.; Alonso, G.; Morrice, N.; Jones, M.; Meier, R.; Cohen, P.; Nebreda, A.R.
EMBO J. 15, 4156-4164, 1996
A:Title: Purification and cDNA cloning of SAPKK3, the major activator of RK/p38 in stress
A:Reference number: S71631; MUID:97015116
A:Accession: S71633
A:Molecule type: mRNA
A:Residues: 1-162 <CUE>
A:Cross-references: EMBL:X8067; NID:g1495709; PID:e245959; PID:g1495710
A:Experimental source: cell type fibroblast
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1996
C:Genetics:
A:Gene: SAPKK3b
A:Superfamily: kinase-related transforming protein; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:1-162/Domain: protein kinase homology (fragment) <KIN>

Query Match 34.8%; Score 48; DB 2; Length 162;
Best Local Similarity 37.5%; Pred. No. 5.2;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 5 PKVGLGKVAEVLRLVLIANKKAS 28
DB 82 PEDILGKIAVSIVKALEHLHLS 105

RESULT 15
C64073
rbs repressor homolog - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 21-Aug-1998
C:Accession: C64073
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350830
A:Accession: C64073
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-332 <TIGR>
A:Cross-references: GB:U32732; GB:L42023; NID:g1573480; PID:g1573487; TIGR:HI0506
C:Superfamily: lac repressor

```

```

Query Match 34.8%; Score 48; DB 2; Length 332;
Best Local Similarity 44.8%; Pred. No. 11;
Matches 13; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 TIYPDKVLGKVAEVLRLVLIANKKASR 29
DB 286 TICOPKAEGLKAVETLLQRIKPNENYR 314

```

Search completed: November 13, 1999, 12:07:55
Job time: 2067 sec

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	138	100.0	416	1	P46_MYCHY	P46192 mycoplasma
2	56	40.6	146	1	RL13 BORBU	051314 borrelia bu
3	55	39.9	397	1	MPK2_CYPCA	Q90321 cyprinus ca
4	52	37.7	209	1	HS2_BACSU	Q34912 bacillus su
5	51	37.0	398	1	MPK2_CHICK	Q90891 gallus gall
6	51	37.0	400	1	MPK2_HUMAN	P36507 homo sapien
7	50	36.2	394	1	MPK1_XENLA	Q05116 xenopus lae
8	49.5	35.9	774	1	DCMA_METJA	Q57617 methanococc
9	49	35.5	212	1	HS2_LACLA	Q02130 lactococcus
10	49	35.5	401	1	MPK2_MOUSE	Q69332 mus musculus
11	49	35.5	400	1	MPK2_RAT	P36506 rattus norv
12	48	34.8	275	1	IF2A_PYRHO	O58655 pyrococcus
13	48	34.8	334	1	MPK6_HUMAN	P52564 homo sapien
14	48	34.8	334	1	MPK6_MOUSE	P70236 mus musculu
15	48	34.8	331	1	RBSR_HAEIN	P44329 haemophilus
16	47	34.1	318	1	MPK3_HUMAN	P46734 homo sapien
17	47	34.1	314	1	MPK3_MOUSE	Q09110 mus musculu
18	47	34.1	177	1	YASO_METJA	Q58450 methanococc
19	46.5	33.7	405	1	HMPD_METJA	O58116 methanococc
20	46	33.3	393	1	MPK1_CRIGR	Q63980 cricetus
21	46	33.3	392	1	MPK1_HUMAN	Q02750 homo sapien
22	46	33.3	392	1	MPK1_MOUSE	P31938 mus musculu
23	46	33.3	392	1	MPK1_RABIT	P29678 oryctolagus
24	46	33.3	392	1	MPK1_RAT	P29678 oryctolagus
25	46	33.3	388	1	MPK1_SERCA	Q01986 rattus norv
26	46	33.3	244	1	T2P5_ECOLI	Q91447 serinus can
27	46	33.3	305	1	Y41F_RHISN	P04390 escherichia
28	45	32.6	487	1	HEP_DROME	P55489 rhizobium s
29	45	32.6	384	1	TPMB_TREPH	Q23977 drosophila
30	44.5	32.2	508	1	TPK2_TREPH	P29720 treponema p
31	44.5	32.2	1627	1	MPK2_CHICK	P32490 saccharomyc
32	44.5	32.2	1612	1	TP2B_CRIL0	Q42131 gallus gall
33	44.5	32.2	1626	1	TP2B_HUMAN	Q64399 cricetus
34	44	31.9	1857	1	FAS2_PENPA	Q02880 homo sapien
35	44	31.9	210	1	HS2M_ARATH	P13568 penicillium
36	44	31.9	329	1	RBSR_ECOLI	Q96331 arabidopsis
37	44	31.9	155	1	RL22_PYRHO	P25551 escherichia
38	44	31.9	561	1	YGGA_YEAST	O59423 pyrococcus
39	43.5	31.5	146	1	RL13_MYCGE	P53166 saccharomyc
40	43.5	31.5	486	1	YG4J_MYCGE	P47657 mycoplasma
41	43.5	31.5	403	1	YMB8_YEAST	P33303 saccharomyc
42	43	31.2	350	1	ARGI_SOYBN	Q04697 saccharomyc
43	43	31.2	397	1	GDN_MOUSE	O49046 glycine max
44	43	31.2	350	1	GDN_MOUSE	O07235 mus musculu

RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., OUCKENBUSH J., SALZBERG S., HANSON M.,
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RL burgdorferi";
RT NATURE 390:580-586(1997).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE EARLY ASSEMBLY PROTEINS OF
CC THE 50S RIBOSOMAL SUBUNIT (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L13P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001140; G2688238; -;
DR PROSITE: PS00783; RIBOSOMAL_L13; 1.
DR TIGR: B80339; -;
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 146 AA; 16693 MW; 73023CE1 CRC32;

Query Match 40.6%; Score 56; DB 1; Length 146;
Best Local Similarity 55.6%; Pred. No. 0.12;
Matches 10; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 6 DKVLGKVAVEVLRLIAK 23
DB 28 DRILGKVAVDVVKILRK 45
RESULT 3
MPK2_CYPCA STANDARD; PRT; 397 AA.
AC Q90321;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE 2
DE (EC 2.7.1.1) (MAP KINASE KINASE 2) (MAPKK 2) (ERK ACTIVATOR KINASE 2)
DE (MAPK/ERK KINASE 2) (MEK2).
GN MKK.
OS CYPRINUS CARPIO (COMMON CARP).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELOSTEI; OSTARIOPHYSI; CYPRINIFORMES; CYPRINOIDEA;
OC CYPRINIDAE; CYPRININAE; CYPRINUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 94146120.
RA HUANG C.-J., LEE M.-S., CHANG G.-D., HUANG F.-L., LO T.-B.;
RT "Molecular cloning and sequencing of a carp cDNA encoding mitogen-
RT activated protein kinase kinase";
RL BIOCHIM. BIOPHYS. ACTA 1220:223-225(1994).
CC -!- FUNCTION: CATALYZES THE CONCOMITANT PHOSPHORYLATION OF A THREONINE
CC AND A TYROSINE RESIDUE IN A THR-GLU-TYR SEQUENCE LOCATED IN MAP
CC KINASES (BY SIMILARITY).
CC -!- PTM: MAPKK IS ITSELF DEPENDENT ON SER/THR PHOSPHORYLATION FOR
CC ACTIVITY CATALYZED BY MAP KINASE KINASES (RAF) (BY
CC SIMILARITY).
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES. BELONGS TO THE MAP KINASE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L23935; G397678; -;
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM: PF00069; pkinase; 2.
DR HSP: P11362; IAGW.
DR TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; TYROSINE-PROTEIN KINASE;
KW ATP-BINDING; PHOSPHORYLATION.
FT DOMAIN 69 366 PROTEIN KINASE.
FT NP_BIND 75 83 ATP (BY SIMILARITY).
FT BINDING 98 98 ATP (BY SIMILARITY).
FT ACT_SITE 191 191 BY SIMILARITY.
FT MOD_RES 219 219 PHOSPHORYLATION (BY RAF) (BY SIMILARITY).
FT MOD_RES 223 223 PHOSPHORYLATION (BY RAF) (BY SIMILARITY).
FT DOMAIN 263 312 PRO-RICH.
SQ SEQUENCE 397 AA; 44147 MW; AF24EDFA CRC32;
Query Match 39.9%; Score 55; DB 1; Length 397;
Best Local Similarity 45.5%; Pred. No. 0.48;
Matches 10; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 5 PKVLGKVAVEVLRLIAKKNK 26
DB 163 PEEILGKVSIAVLRLGLVLRK 184
RESULT 4
HIS2_BACSU STANDARD; PRT; 209 AA.
AC O34912;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHOSPHORIBOSYL-AMP CYCLOHYDROLASE (EC 3.5.4.19) / PHOSPHORIBOSYL-ATP
DE PYROPHOSPHOHYDROLASE (EC 3.6.1.31).
GN HISI OR HISIE.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RA LAZAREVIC V., SOLDÓ B., RIVOLTA C., REYNOLDS S., MAUEL C.,
RA KARAMATA D.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: 5-PHOSPHORIBOSYL-ATP + H(2)O -
CC 5-PHOSPHORIBOSYL-AMP + PYROPHOSPHATE.
CC -!- CATALYTIC ACTIVITY: 5-PHOSPHORIBOSYL-AMP + H(2)O -
CC 5-(5-PHOSPHO-D-RIBOSYLAMINOFORMIMINO)-1-(5-PHOSPHO-RIBOSYL)
CC IMIDAZOLE-4-CARBOXAMIDE.
CC -!- PATHWAY: SECOND AND THIRD STEPS IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF017113; G2618871; -;
DR EMBL: Z99121; E1186174; -;
DR SUBTILIST; BG12603; HISI.
KW HISTIDINE BIOSYNTHESIS; MULTIFUNCTIONAL ENZYME; HYDROLASE.
SQ SEQUENCE 209 AA; 23901 MW; D61A5E7E CRC32;

Query Match 37.7%; Score 52; DB 1; Length 209;
Best Local Similarity 52.4%; Pred. No. 0.7;
Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 6 DKVLGKVAEVLRLIAKKNK 26
||:| || | | :| | ||:
Db 146 DKILKKVGEASEVITIAAKNR 166

RESULT	5
MPK2_CHICK	
ID	MPK2_CHICK STANDARD; PRT: 398 AA.
AC	Q90891;
DT	01-NOV-1997 (REL. 35, CREATED)
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE	DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE
DE	(EC 2.7.1.-) (MAP KINASE 2) (MAPKK 2) (ERK ACTIVATOR KINASE 2
DE	(MACK/ERK KINASE 2) (MEK2).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.

```

[1]
RN SEQUENCE FROM N.A.
RP STRAIN-WHITE LEGHORN; TISSUE=LIVER;
RC WANG H., PINSONNEAULT S., MEURY L., MORAIS R.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: CATALYZES THE CONCOMITANT PHOSPHORYLATION OF A THREONINE
CC AND A TYROSINE RESIDUE IN A THR-GLU-TYR SEQUENCE LOCATED IN MAP
CC KINASES. ACTIVATES THE ERK1 AND ERK2 MAP KINASES (BY SIMILARITY).
CC -!- PTM: MAPKK IS ITSELF DEPENDENT ON SER/THR PHOSPHORYLATION FOR
CC ACTIVITY CATALYZED BY MAP KINASE KINASES (RAF) (BY
CC SIMILARITY).
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES. BELONGS TO THE MAP KINASE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@sib-sib.ch).

```

CC	EMBL	L28703	G994710	PROSITE	PS00107	PROTEIN_KINASE_ATP	1.	PROSITE	PS00108	PROTEIN_KINASE_ST	1.	PROSITE	PS0011	PROTEIN_KINASE_DOM	1.	PFAM	PF00069	pkInase	2.	DR	HSP	P24941	1A01	TRANSFERASE	SERINE/THREONINE-PROTEIN_KINASE	TYROSINE-PROTEIN_KINASE	ATP-BINDING	PHOSPHORYLATION	DOMAIN	70	367	PROTEIN_KINASE	NP_BIND	76	84	ATP (BY SIMILARITY)	FT	BINDING	99	99	ATP (BY SIMILARITY)	FT	ACT_SITE	192	192	BY SIMILARITY	FT	MOD_RES	220	220	PHOSPHORYLATION (BY RAF)	(BY SIMILARITY)	FT	MOD_RES	224	224	PHOSPHORYLATION (BY RAF)	(BY SIMILARITY)	FT	DOMAIN	284	313	PRO-RICH	FT	SEQUENCE	398	AA	44077	NW	1429CSEF2	CRC32	SQ
CC	EMBL	L28703	G994710	PROSITE	PS00107	PROTEIN_KINASE_ATP	1.	PROSITE	PS00108	PROTEIN_KINASE_ST	1.	PROSITE	PS0011	PROTEIN_KINASE_DOM	1.	PFAM	PF00069	pkInase	2.	DR	HSP	P24941	1A01	TRANSFERASE	SERINE/THREONINE-PROTEIN_KINASE	TYROSINE-PROTEIN_KINASE	ATP-BINDING	PHOSPHORYLATION	DOMAIN	70	367	PROTEIN_KINASE	NP_BIND	76	84	ATP (BY SIMILARITY)	FT	BINDING	99	99	ATP (BY SIMILARITY)	FT	ACT_SITE	192	192	BY SIMILARITY	FT	MOD_RES	220	220	PHOSPHORYLATION (BY RAF)	(BY SIMILARITY)	FT	MOD_RES	224	224	PHOSPHORYLATION (BY RAF)	(BY SIMILARITY)	FT	DOMAIN	284	313	PRO-RICH	FT	SEQUENCE	398	AA	44077	NW	1429CSEF2	CRC32	SQ

```
Query Match          37.0%; Score 51; DB 1; Length 398;
Best Local Similarity 45.5%; Pred.'NO. 2;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
```

Qy 5 PDKVLGKVAEVLRLVLI AKNK 26
|:::||||:: ||| | : |
Db 164 PEEILGKVSIAVLRLGLAYLREK 185

RESULT 6

```

MPK2_HUMAN
ID MPK2_HUMAN STANDARD; PRT; 400 AA.
AC P36507;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE 2
DE (EC 2.7.1.-) (MAP KINASE KINASE 2) (MAPKK 2) (ERK ACTIVATOR KINASE 2)
DE DE (MAPK/ERK KINASE 2) (MEK2).
DE GN PRMK2 OR MEK2 OR MKK2.
DE OS HOMO SAPIENS (HUMAN).
DE OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
DE PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
DE [1]
DE RP SEQUENCE FROM N.A.
DE RX MEDLINE; G3266604.
DE RA ZHENG C.-F.; GUAN K.-L.;
DE RT "Cloning and characterization of two distinct human extracellular
DE RT signal-regulated kinase activator kinases, MEK1 and MEK2.";
DE RL J. BIOL. CHEM. 268:11435-11439(1993).
DE CC -1- FUNCTION: CATALYZES THE CONCOMITANT PHOSPHORYLATION OF A THREONINE
DE CC AND A TYROSINE RESIDUE IN A THR-GLU-TYR SEQUENCE LOCATED IN MAP
DE CC KINASES. ACTIVATES THE ERK1 AND ERK2 MAP KINASES (BY SIMILARITY).
DE CC -1- PTM: MAPKK IS ITSELF DEPENDENT ON SER/THR PHOSPHORYLATION FOR
DE CC ACTIVITY CATALYZED BY MAP KINASE KINASES (RAF OR MEKK1).
DE CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
DE CC PROTEIN KINASES. BELONGS TO THE MAP KINASE KINASE FAMILY.
DE -----
DE CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
DE CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
DE CC the European Bioinformatics Institute. There are no restrictions on its
DE CC use by non-profit institutions as long as its content is in no way
DE CC modified and this statement is not removed. Usage by and for commercial
DE CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
DE CC or send an email to license@sib-sib.ch).

```

CC	EMBL; L11285; -; NOT_ANNOTATED_CDS.	
DR	PIR; A46723; A46723.	
DR	MMX; 601263.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.	
DR	PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.	
DR	PFAM; PF00069; pkinase; 1.	
DR	HSSP; P11362; IAGW.	
DR	TRANSFERASE; SERINE/THREONINE-PROTEIN_KINASE; TYROSINE-PROTEIN_KINASE.	
KW	ATP-BINDING; PHOSPHORYLATION.	
KW	DOMAIN 72 369	PROTEIN_KINASE.
FT	NP_BIND 78 86	ATP (BY SIMILARITY).
FT	BINDING 101 101	ATP (BY SIMILARITY).
FT	ACT_SITE 194 194	BY SIMILARITY.
FT	MOD_RES 222 222	PHOSPHORYLATION (BY RAF)
FT	MOD_RES 226 226	PHOSPHORYLATION (BY RAF)
FT	DOMAIN 266 315	PRO-RICH.
SQ	SEQUENCE 400 AA; 444424 MW; 07113E82 CRC32;	

```

Query Match      37.0%; Score 51; DB 1; Length 400;
Best Local Similarity 45.5%; Pred. NO. 2;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

```

Qy 5 PDKVLGKVAVEVLRVLI AKKNK 26
|:::|::|::|::|::|
Db 166 PEEILGKVSIAVLRGLAYLREK 187

RESULT 7
MPK1_XENLA
ID MPK1_XENLA
STANDARD; PRT; 394 AA.

AC	Q05116:	
DT	01-JUN-1994	(REL. 29, CREATED)
DT	01-JUN-1994	(REL. 29, LAST SEQUENCE UPDATE)
DT	15-DEC-1998	(REL. 37, LAST ANNOTATION UPDATE)
DE	DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE 1	

(EC 2.7.1.1-) (MAP KINASE KINASE 1) (MAPKK 1) (ERK ACTIVATOR KINASE 1)
DE (MAPK/ERK KINASE 1) (MEK1).
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATACHIA; PIPODEA; PIPIDEA; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 93178455.
RA KOSAKO H., NISHIDA E., GOTOH Y.;
RT "CDNA cloning of MAP kinase kinase reveals kinase cascade pathways in
RT yeasts to vertebrates";
RL EMBO J. 12:787-794(1993).
CC -!- FUNCTION: CATALYZES THE CONCOMITANT PHOSPHORYLATION OF A THREONINE
CC AND A TYROSINE RESIDUE IN A THR-GLU-TYR SEQUENCE LOCATED IN MAP
CC KINASES.
CC -!- TISSUE SPECIFICITY: IS EXPRESSED IN THE CENTRAL NERVOUS SYSTEM,
CC KIDNEYS, LIVER, INTESTINE AND THE HEMATOPOIETIC SYSTEM.
CC -!- PTM: MAPKK IS ITSELF DEPENDENT ON SER/THR PHOSPHORYLATION FOR
CC ACTIVITY CATALYZED BY MAP KINASE KINASE KINASES (RAF OR MOS).
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES. BELONGS TO THE MAP KINASE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D13700; G222965; -
DR PIR: S36186; S36186.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM: PF00069; pkinase; 1.
DR HSP: Q63450; 1A06.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; TYROSINE-PROTEIN KINASE;
KW ATP-BINDING; PHOSPHORYLATION.
FT INIT MET 0
FT DOMAIN 67 362 PROTEIN KINASE.
FT NP_BIND 73 81 ATP (BY SIMILARITY).
FT BINDING 96 96 ATP (BY SIMILARITY).
FT ACT_SITE 189 189 BY SIMILARITY.
FT MOD_RES 217 217 PHOSPHORYLATION (BY RAF) (BY SIMILARITY).
FT MOD_RES 221 221 PHOSPHORYLATION (BY RAF) (BY SIMILARITY).
FT DOMAIN 261 308 PRO-RICH.
SQ SEQUENCE 394 AA; 43611 MW; 2CEA7591 CRC32;

Query Match 36.2%; Score 50; DB 1; Length 394;
Best Local Similarity 40.9%; Pred. No. 2.8;
Matches 9; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
OY 5 PKVLGKVAEVLRLVIAKKNK 26
I:|||||:|:|:|
Db 161 PEKILGKVSIAVIGITLYREK 182

RESULT 8
DCMA_METJA STANDARD; PRT; 774 AA.
AC Q57617;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CARBON MONOXIDE DEHYDROGENASE ALPHA SUBUNIT (EC 1.2.99.2).
GN CDHA OR MJ0153.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
OC METHANOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.

STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE: 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBEER K., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA SCOTT J.L., GEORHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RL SCIENCE 273:1058-1073(1996).
CC -!- FUNCTION: CATALYZES THE CLEAVAGE OF ACETYL-COA (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: CO + H(2)O + ACCEPTOR = CO(2) + REDUCED
CC ACCEPTOR.
CC -!- COFACTOR: TWO NICKEL AND FOUR IRON-SULFUR CLUSTERS PER ENZYME
CC (PROBABLY TWO 4FE-4S AND TWO 6FE-6S).
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67472; G2826254; -
DR TIGR: MJ0153; -
DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
DR PFAM: PF00037; fer4; 1.
KW OXIDOREDUCTASE; NICKEL; IRON-SULFUR; 4FE-4S.
FT METAL 76 76
FT METAL 79 79 IRON-SULFUR (6FE-6S) (BY SIMILARITY).
FT METAL 94 94 IRON-SULFUR (6FE-6S) (BY SIMILARITY).
FT METAL 408 408 IRON-SULFUR (BY SIMILARITY).
FT METAL 411 411 IRON-SULFUR (BY SIMILARITY).
FT METAL 414 414 IRON-SULFUR (BY SIMILARITY).
FT METAL 418 418 IRON-SULFUR (BY SIMILARITY).
FT METAL 446 446 IRON-SULFUR (BY SIMILARITY).
FT METAL 449 449 IRON-SULFUR (BY SIMILARITY).
FT METAL 452 452 IRON-SULFUR (BY SIMILARITY).
FT METAL 456 456 IRON-SULFUR (BY SIMILARITY).
SQ SEQUENCE 774 AA; 85900 MW; E932E418 CRC32;

Query Match 35.9%; Score 49.5; DB 1; Length 774;
Best Local Similarity 52.4%; Pred. No. 6.8;
Matches 11; Conservative 6; Mismatches 3; Indels 1; Gaps 1;
OY 6 DKVLGKVAEVLRLVIAKKNK 26
I:|||||:|:|:|
Db 371 EKKVGKVAEVAKI-VAKERK 390

RESULT 9
HIS2_LACLA STANDARD; PRT; 212 AA.
AC Q02130;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHOSPHORIBOSYL-AMP CYCLOHYDROLASE (EC 3.5.4.19) / PHOSPHORIBOSYL-ATP
DE PYROPHOSPHORYLASE (EC 3.6.1.31).
GN HIS1 OR HISIE.
OS LACTOCOCCUS LACTIS (SUBSP. LACTIS) (STREPTOCOCCUS LACTIS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC LACTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-NCDO 2118;
 RX MEDLINE; 93015709.
 RA DELORE C.C., EHRLICH S.D., RENAULT P.P.;
 RT "Histidine biosynthesis genes in Lactococcus lactis subsp. lactis.";
 RL J. BACTERIOL. 174:6571-6579(1992).
 CC -1- CATALYTIC ACTIVITY: 5-PHOSPHORIBOSYL-ATP + H(2)O =
 CC 5-PHOSPHORIBOSYL-AMP + PYROPHOSPHATE.
 CC -1- CATALYTIC ACTIVITY: 5-PHOSPHORIBOSYL-AMP + H(2)O =
 CC 5-(5-PHOSPHO-D-RIBOSYLAMINOFORMIMINO)-1-(5-PHOSPHO-RIBOSYL)
 CC IMIDAZOLE-4-CARBOXAMIDE.
 CC -1- PATHWAY: SECOND AND THIRD STEPS IN HISTIDINE BIOSYNTHETIC PATHWAY.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U92974; G2565148; -
 DR PIR; C47754; C47754.
 KW HISTIDINE BIOSYNTHESIS; MULTIFUNCTIONAL ENZYME; HYDROLASE.
 SQ SEQUENCE 212 AA; 24675 MW; 6293D21F CRC32;

Query Match 35.5%; Score 49; DB 1; Length 212;
 Best Local Similarity 60.0%; Pred. No. 2.1;
 Matches 12; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 6 DKVLGKVAVEVLRLVIAKKN 25
 ||||| I I I I I
 DB 139 DKVLKVGEEASEVTIAKKN 158

RESULT 10
 ID MPK2_MOUSE STANDARD; PRT; 401 AA.
 AC Q63932;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE 2
 DE (EC 2.7.1.1) (MAP KINASE KINASE 2) (MAPKK 2) (ERK ACTIVATOR KINASE 2)
 DE (MAPK/ERK KINASE 2) (MEK2).
 GN PRKMK2 OR MEK2 OR MKK2.
 OS MUS MUSCULUS (MOUSE)
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.
 RC STRAIN-BALB/C; TISSUE-NEONATAL BRAIN;
 RX MEDLINE; 94128609.
 RA BROTT B.K., ALESSANDRINI A., LARGASPADA D.A., COPELAND N.G.,
 RA JENKINS N.A., CREWS C.M., ERIKSON R.L.;
 RT "MEK2 is a kinase related to MEK1 and is differentially expressed in
 RT murine tissues.";
 RL CELL GROWTH DIFFER. 4:921-929(1993).
 RN [2]
 RP SEQUENCE OF 118-140 AND 210-238.
 RC TISSUE-T-CELL;
 RX MEDLINE; 92390415.
 RA CREWS C.M., ERIKSON R.L.;
 RT "Purification of a murine protein-tyrosine/threonine kinase that
 RT phosphorylates and activates the Erk-1 gene product: relationship to
 RT the fission yeast *byr1* gene product.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:8205-8209(1992).
 CC -1- FUNCTION: CATALYZES THE CONCOMITANT PHOSPHORYLATION OF A THREONINE
 CC AND A TYROSINE RESIDUE IN A THR-GLU-TYR SEQUENCE LOCATED IN MAP
 CC KINASES. ACTIVATES THE ERK1 AND ERK2 MAP KINASES.
 CC -1- ENZYME REGULATION: INHIBITED BY SERINE/THREONINE PHOSPHATASE 2A.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ADULT INTESTINE, KIDNEY, LIVER,

CC LUNG, PANCREAS, SPLEEN, THYMUS, AND AT HIGH LEVELS IN THE NEONATAL
 CC BRAIN. LOWER EXPRESSION IS FOUND IN ADULT BRAIN AND HEART.
 CC -1- PTM: MAPKK IS ITSELF DEPENDENT ON SER/THR PHOSPHORYLATION FOR
 CC ACTIVITY CATALYZED BY MAP KINASE KINASES (RAF OR MEKK1).
 CC LOW LEVELS OF AUTOPHOSPHORYLATION HAVE BEEN OBSERVED.
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
 CC PROTEIN KINASES. BELONGS TO THE MAP KINASE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; S68267; G545203; -
 DR MGD; MGI:101840; PRKMK2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
 DR PFAM; PF00069; pkinase; 1.
 DR HSP; P24941; 1A01.
 DR TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; TYROSINE-PROTEIN KINASE;
 KW ATP-BINDING; PHOSPHORYLATION.
 FT DOMAIN 72 370 PROTEIN KINASE.
 FT NP_BIND 78 86 ATP (BY SIMILARITY).
 FT BINDING 101 101 ATP.
 FT ACT_SITE 194 194 BY SIMILARITY.
 FT MOD_RES 222 222 PHOSPHORYLATION (BY RAF) (BY SIMILARITY).
 FT MOD_RES 226 226 PHOSPHORYLATION (BY RAF) (BY SIMILARITY).
 FT MUTAGEN 101 101 K->M; INACTIVATION.
 FT DOMAIN 266 316 PRO-RICH.
 FT SEQUENCE 401 AA; 44436 MW; 5BDCB22 CRC32;

Query Match 35.5%; Score 49; DB 1; Length 401;
 Best Local Similarity 45.5%; Pred. No. 4;
 Matches 10; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLVIAKKN 26
 | : ||||| : | : |
 DB 166 PEDILGKVSIAVLRLGLAYLREK 187

RESULT 11
 ID MPK2_RAT STANDARD; PRT; 400 AA.
 AC P36506;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE 2
 DE (EC 2.7.1.1) (MAP KINASE KINASE 2) (MAPKK 2) (ERK ACTIVATOR KINASE 2)
 DE (MAPK/ERK KINASE 2) (MEK2).
 GN PRKMK2 OR MEK2 OR MKK2.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93215844.
 RA OTSU M., TERADA Y., OKAYAMA H.;
 RT "Isolation of two members of the rat MAP kinase kinase gene family.";
 RL FEBS LETT. 320:246-250(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93330248.
 RA WU J., HARRISON J.K., DENT P., LYNCH K.R., WEBER M.J.,
 RA STURGILL T.W.;
 RT "Identification and characterization of a new mammalian mitogen-
 RT activated protein kinase kinase, MKK2.";
 RL MOL. CELL. BIOL. 13:4539-4548(1993).

CC (BY SIMILARITY).
CC -1- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA
CC CHAIN (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP000004; D1031001; -
DR INITIATION FACTOR; PROTEIN BIOSYNTHESIS. KW
SQ SEQUENCE 275 AA; 31980 MW; 1206D3E3 CRC32;
CC

```

Query Match      34.8%; Score 48; DB 1; Length 275;
Best Local Similarity 34.3%; Pred. No. 3.9;
Matches 12; Conservative 7; Mismatches 8; Indels 8; Gaps 1;

QY      3 YKPKVILGKVAVEVLRL-----IAKNKASR 29
      || :||| :||| :||| :||| :||| :|||
Db      234 YKAEVLEDAEILRLVQAGGEATILLRKRIR 268

RESULT 13
MPK6 HUMAN
ID      MPK6_HUMAN      STANDARD;      PRT;      334 AA.
AC      P52564;
DT      01-OCT-1996 (REL. 34, CREATED)
DT      01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT      01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE      DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE 6
DE      (EC 2.7.1.1-) (MAP KINASE KINASE 6) (MAPKK 6) (MAPKK/ERK KINASE 6)
DE      (SAPKK3)
GN      PRMK6 OR MKK6 OR MEK6.

```

[illegible]

MOL. CELL. BIOL. 16:1247-1255(1996).
 [2]
 RN SEQUENCE FROM N.A.
 RN RP
 RC TISSUE-T-CELL;
 RX MEDLINE; 96212215.
 RA STEIN B., BRADY H., YANG M.X., YOUNG D.B., BARBOSA M.S.;
 RT "Cloning and characterization of MEK6, a novel member of the mitogen-
 RT activated protein kinase kinase cascade.";
 RL J. BIOL. CHEM. 271:11427-11433(1996).
 [3]
 RN SEQUENCE FROM N.A., AND MUTAGENESIS (SHORT FORM).
 RN RP
 RC TISSUE-PLACENTA;
 RX MEDLINE; 96216353.
 RA HAN J., LEE J.-D., JIANG Y., LI Z., FENG L., ULEVITCH R.J.;
 RT "Characterization of the structure and function of a novel MAP kinase
 RT kinase (MKK6).";
 RL J. BIOL. CHEM. 271:2886-2891(1996).
 [4]
 RN SEQUENCE FROM N.A.
 RN RP
 RC MEDLINE; 96278799.
 RX MORIGUCHI T., KUROYANAGI N., YAMAGUCHI K., GOTOH Y., IRIE K., KANO T.,
 RA SHIRAKABE K., MURO Y., SHIBUYA H., MATSUMOTO K., NISHIDA E.,
 RA HAGIWARA M.;
 RT "A novel kinase cascade mediated by mitogen-activated protein kinase
 RT kinase 6 and MKK3.";
 RL J. BIOL. CHEM. 271:13675-13679(1996).

```

RESULT 14
MPK6_MOUSE STANDARD; PRT; 334 AA.
AC P70236;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 13-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE 6
DE (SAPKK3).
DE DE 2.7.1.-) (MAP KINASE KINASE 6) (MAPKK 6) (MAPK/ERK KINASE 6)
DE DE
GN PRKM6 OR SAPKK3.
OS MUS MUSCULUS (MURAE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]
SEQUENCE FROM N.A.
MEDLINE: 97015116.
RA CUENDA A., ALONSO G., MORRICE N., JONES M., MEIER R., COHEN P.,
RA "BERGDA A.R.;
RA Purification and cDNA cloning of SAPKK3, the major activator of
RT RK/p38 in stress- and cytokine-stimulated monocytes and epithelial
RT cells."
RL EMBO J. 15:4156-4164(1996).
CC -I- FUNCTION: CATALYZES THE CONCOMITANT PHOSPHORYLATION OF A THREONINE
AND A TYROSINE RESIDUE IN MAP KINASE P38 EXCLUSIVELY.
CC -I- ENZYME REGULATION: PROBABLY ACTIVATED BY DUAL PHOSPHORYLATION ON
SER-207 AND THR-211 (BY SIMILARITY).
CC -I- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
PROTEIN KINASES. BELONGS TO THE MAP KINASE KINASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: X97052; E244590; -.
MGD: MGI:108032; PRKM6.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; P24941; IAKI.
DR TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; TYROSINE-PROTEIN KINASE;
KW ATP-BINDING; PHOSPHORYLATION.
FT DOMAIN 53 314 PROTEIN KINASE.
FT NP_BIND 59 67 ATP (BY SIMILARITY).
FT BINDING 82 82 ATP (BY SIMILARITY).
FT ACT_SITE 179 179 BY SIMILARITY.
FT MOD_RES 207 207 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 211 211 PHOSPHORYLATION (PROBABLE)..
SQ SEQUENCE 334 AA; 37432 MW; 08C9FE40 CRC32;

Query Match 34.8%; Score 48; DB 1; Length 334;
Best Local Similarity 37.5%; Pred.No. 4.7;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 5 PDKVLGVKVAVEVLRVLIAKNKAS 28
| : |||| :: | : |
Db 151 PEDILGLIAVSIVKALEHLHLSKL 174

RESULT 15
RBSR_HAEN ID RBSR_HAEN STANDARD; PRT; 331 AA.
AC P44329;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE RIBOSE OPERON REPRESSOR.
```

RESULT 15				
RBSR_HAEIN				
ID	RBSR_HAEIN	STANDARD;	PRT;	331 AA.
AC	P44329;			
DT	01-NOV-1995	(REL. 32, CREATED)		
DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)		
DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)		
DE	RIBOSE OPERON REPRESSOR.			

Search completed: November 13, 1999, 10:33:31
Job time: 5181 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:22 ; Search time 139.86 Seconds
(without alignments)
12.761 Million cell updates/sec

Title: US-08-913-430-4

Perfect score: 138

Sequence: 1 TTYKPKVLGKVAVEVLRLVLIAKKNKASR 29

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPTREMBL10.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	37.0	96	1 027344	027344 methanobact
2	48	34.8	429	2 025470	025470 helicobacte
3	48	34.8	419	4 014733	014733 homo sapien
4	48	34.8	401	4 014648	014648 homo sapien
5	48	34.8	419	4 060452	060452 homo sapien
6	48	34.8	462	4 060453	060453 homo sapien
7	48	34.8	419	4 014816	014816 homo sapien
8	48	34.8	785	4 094782	094782 homo sapien
9	48	34.8	660	5 000885	000885 dictyosteli
10	48	34.8	346	11 035406	035406 mus musculu
11	48	34.8	237	11 060521	060521 mus musculu
12	48	34.8	162	11 070273	070273 mus musculu
13	48	34.8	389	11 035720	035720 mus musculu
14	48	34.8	391	11 035872	035872 mus musculu
15	48	34.8	468	11 034780	034780 mus musculu
16	48	34.8	295	11 070242	070242 mus musculu
17	48	34.8	311	11 070243	070243 mus musculu
18	48	34.8	419	11 035871	035871 mus musculu
19	47	34.1	347	4 099441	099441 homo sapien
20	47	34.1	1026	5 091494	091494 caenorhabdi
21	47	34.1	505	5 021307	021307 caenorhabdi
22	47	34.1	347	11 097293	097293 mus musculu
23	46	33.3	1045	2 085234	085234 lactobacill
24	46	33.3	429	2 092162	092162 helicobacte
25	46	33.3	408	3 093924	093924 fusarium mo
26	46	33.3	539	10 080918	080918 arabidopsis
27	45	32.6	167	2 033824	033824 thriobacilli
28	45	32.6	492	5 018411	018411 drosophila
29	44	31.9	790	2 050638	050638 mycobacteri

ALIGNMENTS

RESULT 1

027344 ID 027344 PRELIMINARY; PRT; 96 AA.

AC 027344;
DT 01-JAN-1998 (TREMELrel. 05, Created)
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMELrel. 07, Last annotation update)
DE PHOSPHORIBOSYL-AMP CYCLOHYDROLASE HOMOLOG.

GN MTH283.

OS Methanobacterium thermoautotrophicum.

OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;

OC Methanobacterium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-DELTA H;

RX MEDLINE; 98037514.

RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,

RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,

RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,

RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,

RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,

RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,

RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;

RT "Complete genome sequence of Methanobacterium thermoautotrophicum

deltaH: functional analysis and comparative genomics.";

RL J. Bacteriol. 179:7135-7155(1997).

DR EMBL; AF0000893; AAB85765.1; -.

KW Hydrolase.

SQ SEQUENCE 96 AA; 11122 MW; AC721C92 CRC32;

Query Match 37.0%; Score 51; DB 1; Length 96;

Best Local Similarity 50.0%; Pred. No. 2.8;

Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 6 DKVLGKVAVEVLRLVLIAKKN 25

Db 39 DKILEKIGEEAAEVIATSKN 58

RESULT 2

025470 ID 025470 PRELIMINARY; PRT; 429 AA.

AC 025470;

DT 01-JAN-1998 (TREMELrel. 05, Created)

DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)

DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)

DE HYPOTHETICAL 49.4 KD PROTEIN.

GN HP0781.

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

8
RESULT
O94782

```

Qy      5 PDKVLGKVAEVLRLVI 21
      |: ||||:| :||: |:
Db      386 PEPVLGKIAFQVLQGLV 402

RESULT 10
O35406 ID 4035406 PRELIMINARY; PRT; 346 AA.
AC AC O35406;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE DE PROTEIN KINASE, MITOGEN ACTIVATED, KINASE 7 (MKK7).

```

RP	SEQUENCE FROM N.A.
RX	MEDLINE; 97015116.
RA	CUENDA A., ALONSO G., MORRICE N., JONES M., MEIER R., COHEN P., NEBREDA A.R.;
RA	NEBREDA A.R.;
RT	"Purification and cDNA cloning of SAPKK3, the major activator of RK/p38 in stress- and cytokine-stimulated monocytes and epithelial cells.";
RT	EMBO J. 15:4156-4164(1996).
RL	EMBL; X98067; CAA66670.1; "-
DR	MGL; MGI:108032; PRKMR6.
DR	PFAM; PF00069; pkinase. 1.
DR	NON_TER 162
FT	FT 162
SQ	SEQUENCE 162 AA; 18200 MW; 812DB422 CRC32; Query Match 34.8%; Score 48; DB 11; Length 162; Best Local Similarity 37.5%; Pred. No. 13; Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY	5	PDRVGVAVAVEVRLVLLAKNNAS	28
		: : : : : :	
Db	82	PEDILGKIAVSIVKALEHLSKLS	105
RESULT	13		
O35720			
ID	O35720	PRELIMINARY;	PRT; 389 AA.
AC	O35720;		
DT	01-JAN-1998	(TREMBlrel. 05, Created)	
DT	01-JAN-1998	(TREMBlrel. 05, Last sequence update)	
DT	01-NOV-1998	(TREMBlrel. 08, Last annotation update)	
DE	DE	PROTEIN KINASE, MITOGEN ACTIVATED, KINASE 7 (MAP KINASE KINASE 7)	
DE	DE	(FRAGMENT).	
GN	PRKMk7.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CD-1; TISSUE-TESTIS;		
RX	MEDLINE; 97352799.		
RA	TOURNIER C., WHITMARSH A.J., CAVANAGH J., BARRETT T., DAVIS R.J.;		
RT	"Mitogen-activated protein kinase 7 is an activator of the		
RT	c-Jun NH2-terminal kinase";		
RL	Proc. Natl. Acad. Sci. U.S.A. 94:7337-7342(1997).		
DR	EMBL; U93031; AAB63448.1; -.		
DR	MGD; MGI:1099437; PRKMk7.		
DR	PFAM; PF00069; pkinase.1.		

```

SQ SEQUENCE 389 AA; 43798 MW; DACA7D75 CRC32;

Query Match 34.8%; Score 48; DB 11; Length 389;
Best Local Similarity 36.4%; Pred. No. 33;
Matches 8; Conservative 7; Mismatches 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLIAKKNK 26
|:::|::|::|::|::|::|
Db 185 PERILGRMTVAIVRALYLYLEK 206

RESULT 14
O35872 PRELIMINARY; PRT; 391 AA.
ID O35872 AC O35872:
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE PROTEIN KINASE, MITOGEN ACTIVATED, KINASE 7 (MKK7B).
GN PKMK7 OR MKK7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

```

●●●●●

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97460085.
 RA HOLLAND P.M., MAGALI S., CAMPBELL J.S., NOSELLI S., COOPER J.A.;
 RT "MKK7 is A stress-activated mitogen-activated protein kinase
 functionally related to hemipterous";
 RL J. Biol. Chem. 272:24994-24998(1997).
 RR EMBL: U74464; AAC53365.1; -.
 DR MGD; MGI:1099437; PRKMK7.
 DR PFAM; PF00069; pkinase; 1.
 SQ SEQUENCE 391 AA; 44210 MW; 845D093C CRC32;

Query Match 34.8%; Score 48; DB 11; Length 391;
 Best Local Similarity 36.4%; Pred. No. 34;
 Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLVLIANKK 26
 |:::|:::|:::|:::|:::|
 Db 187 PERILGKMTVAIVKALYILKEK 208

RESULT 15
 O54780 PRELIMINARY; PRT; 468 AA.
 AC O54780;
 DT 01-JUN-1998 (TREMELrel. 06, Created)
 DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
 DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)
 DE PROTEIN KINASE, MITOGEN ACTIVATED, KINASE 7
 DE (MITOGEN-ACTIVATED PROTEIN KINASE 7).
 GN PRKMK7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98046034.
 RA MORIGUCHI T., TOYOSHIMA F., MASUYAMA N., HANAFUSA H., GOTOH Y.,
 RA NISHIDA E.;
 RT "A novel SAPK/JNK kinase, MKK7, stimulated by TNFalpha and cellular
 stresses";
 RL EMBO J. 16:7045-7053(1997).
 DR EMBL; AB005654; BAA24383.1; -.
 DR MGD; MGI:1099437; PRKMK7.
 DR PFAM; PF00069; pkinase; 1.
 SQ SEQUENCE 468 AA; 52496 MW; A644CCBE CRC32;

Query Match 34.8%; Score 48; DB 11; Length 468;
 Best Local Similarity 36.4%; Pred. No. 41;
 Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 PDKVLGKVAVEVLRLVLIANKK 26
 |:::|:::|:::|:::|:::|
 Db 231 PERILGKMTVAIVKALYILKEK 252

Search completed: November 13, 1999, 12:55:24
 Job time: 3023 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:59:46 ; Search time 104.22 Seconds
(without alignments)
3.636 Million cell updates/sec

Title: US-08-913-430-5
Perfect score: 78
Sequence: 1 AEQAITKLEGFDTQ 16

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	78	100.0	281	1 R06279	Swine enzootic pne
2	78	100.0	419	1 R21829	Sequence of surfac
3	78	100.0	16	1 W01035	Mycoplasma 46-48 k
4	78	100.0	419	1 W01037	Mycoplasma 46-48 k
5	41	52.6	1456	1 R70419	Rat homologue of h
6	41	52.6	1411	1 R72343	Wilson disease pro
7	38	48.7	3567	1 R44431	eryA region polype
8	38	48.7	257	1 W80609	S. pneumoniae prot
9	37	47.4	72	1 R71345	Carboxy terminal b
10	37	47.4	847	1 R88320	IL-4 Stat peptide.
11	37	47.4	432	1 W21894	Asparaginyl-trna s
12	37	47.4	432	1 W63842	S. aureus asparagi
13	37	47.4	430	1 W63843	S. aureus asparagi
14	37	47.4	737	1 Y05221	Human Stat6b prote
15	37	47.4	819	1 Y05222	Human Stat6c prote
16	37	47.4	848	1 Y05223	Human Stat6 protei
17	36	46.2	2101	1 R47173	Sequence of the in
18	36	46.2	2272	1 W21731	GAI4/HA/NUMA fusio
19	36	46.2	2192	1 W21732	LexA/NUMA fusion p
20	36	46.2	519	1 W41158	Biological clock c
21	35	44.9	323	1 R56282	Human tissue plas
22	35	44.9	249	1 R86719	Human proteasome a
23	35	44.9	323	1 R80944	Receptor-associate
24	35	44.9	981	1 W46499	Amino acid sequenc
25	35	44.9	136	1 W57416	Isopeptidase enzym
26	35	44.9	189	1 W82021	Bovine partial PA2
27	35	44.9	493	1 W97823	Trichoderma reesei
28	34.5	44.2	865	1 P70175	Sequence encoded b
29	34	43.6	527	1 R14183	Product of clone I
30	34	43.6	527	1 R38099	Schistosoma mans
31	34	43.6	3054	1 R40841	Translation of TEV
32	34	43.6	526	1 R17735	Heliothis virescen
33	34	43.6	238	1 W80542	S. pneumoniae prot
34	34	43.6	772	1 Y01069	Human l(3)mbt prot
35	34	43.6	738	1 Y01070	Human l(3)mbt prot
36	34	43.6	826	1 Y01071	Mouse l(3)mbt prot
37	33	42.3	295	1 R05562	Laminin-binding p
38	33	42.3	253	1 P70569	Sequence of lamini
39	33	42.3	370	1 R12957	Asn(72), Lys(73) C
40	33	42.3	577	1 W06554	R. capsulatus acet
41	33	42.3	429	1 W20678	H. pylori transpor
42	33	42.3	3059	1 W19696	ATM mutant 5319ins
43	33	42.3	3055	1 W19697	ATM mutant 5435del

44 33 42.3 3056 1 W19698 ATM mutant E2904G.
45 33 42.3 967 1 W78189 Human secreted pro

ALIGNMENTS

RESULT 1

R06279 ID R06279 standard; protein; 281 AA.
AC R06279;
DT 19-DEC-1990 (first entry)
DE Swine enzootic pneumonia mycoplasma surface antigen polypeptide.
KW Mycoplasma pneumoniae; enzootic pneumoniae; ds.
OS Mycoplasma hyopneumoniae.
PN J02167079-A.
PD 27-JUN-1990.
PF 21-DEC-1988; 322829.
PR 21-DEC-1988; JP-322829.
PA (NISE-) NIPPON SEIFUN KK.
DR WPI; 90-241949/32.
DR N-PSDB; Q05578.
PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used
PT for early detection and treatment of mycoplasma or enzootic
PT pneumonia of pigs
PS Claim 1; Page 570; 28pp; Japanese.
CC Sequence encoding polypeptides may be used as probes for the early
CC diagnosis of pneumonia in pigs. The sequences may be used to produce
CC an expression vector and transform a suitable host.
SQ Sequence 281 AA;

Query Match 100.0%; Score 78; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEQAITKLEGFDTQ 16

DB 163 AEQAITKLEGFDTQ 178

RESULT 2

R21829 ID R21829 standard; protein; 419 AA.
AC R21829;
DT 02-NOV-1992 (first entry)
DE Sequence of surface antigen 46kd.
KW Swine pneumonia; epidemic; diagnosis; therapy.
OS Mycoplasma hyopneumoniae.
PN EP-475185-A.
PD 18-MAR-1992.
PF 27-AUG-1991; 114335.
PR 27-AUG-1990; JP-224945.
PA (NIFL-) NIPPON FLOUR MILLS.
PI Seto Y, Futo S, Mitsuse S, Matsuo K, Tsuna M;
DR WPI; 92-089874/12.
DR N-PSDB; Q22042.
PT DNA and peptide of mycoplasma hyopneumoniae - useful for
PT diagnosis and treatment of swine mycoplasma pneumonia
PS Disclosure; Page 4-6 and pages 34-36; 45pp; English.
CC The inventors claim DNA encoding a surface antigen and primers used
CC in a method of diagnosing mycoplasma pneumonia of swine.
CC Mycoplasma cells were collected from the culture of M. hp. lysed in a
CC purification thereof. The DNA of M. hp. is cleaved with a restriction
CC enzyme HindIII, the resulting fragment is inserted into a plasmid
CC pUC19 at the site cleaved with HindIII. This plasmid is transformed
CC into E. coli as a host cell, colony hybridization is performed by the
CC DNA probe method utilizing EcoRI fragments of pKUM1 and pKUM2 to
CC select the bacterial cells containing a plasmid (pURR126) carrying the
CC pKUM1 or pKUM2 is deposited with FRI under accession No. FERM P-10318
CC or P-10319. The base sequence of the 46 kd antigen gene is given in

CC Q22042.
SQ Sequence 419 AA;

Query Match 100.0%; Score 78; DB 1; Length 419;

Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEQAITKLEGFDTQ 16

DB 301 AEQAITKLEGFDTQ 316

RESULT 3

W01035
ID W01035 standard; Peptide; 16 AA.

AC W01035; 1997 (first entry)

DE Mycoplasma 46-48 kDa protective antigen fragment CNBR F2.

KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumoniae;
diagnosis; antibody.

OS Mycoplasma hyopneumoniae strain Beaufort.

PN W09628472-A1.

PD 19-SEP-1996.

PF 15-MAR-1996; AU0149.

PR 16-MAR-1995; AU-001789.

PA (UYME) UNIV MELBOURNE.

PI Doughty SW, Lee R, Walker J;

DR WPI; 96-433763/43.

PT Putative protective antigens against Mycoplasma - used for the

detection, prevention or treatment of Mycoplasma infections, esp. M.

PS Claim 13; Page 28; 43pp; English.

CC A 46-48 kDa putative protective antigen against Mycoplasma contains

the N-terminal sequence given in W01034-36. The antigen was isolated from

fragments given in W01034-36. The antigen was isolated from

Mycoplasma hyopneumoniae cells using antibody probes enriched with

CC Mycoplasma-specific antibodies. Other protective antigens were

also identified (see also W01024-32). A gene (T38241) coding for

the 48 kDa antigen (W01037) was isolated from a genomic library.

Protective antigens and antibodies can be used in vaccines for

preventing or treating mycoplasma infections, partic. M.

CC hyopneumoniae infections in swine. They can also be used for

diagnosis.

CC Sequence 16 AA;

Query Match 100.0%; Score 78; DB 1; Length 16;

Best Local Similarity 100.0%; Pred. No. 3.7e-08;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEQAITKLEGFDTQ 16

DB 1 AEQAITKLEGFDTQ 16

RESULT 4

W01037

ID W01037 standard; Protein; 419 AA.

AC W01037; 1997 (first entry)

DE Mycoplasma 46-48 kDa protective antigen.

KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumoniae;

diagnosis; antibody.

OS Mycoplasma hyopneumoniae strain Beaufort.

PN W09628472-A1.

PD 19-SEP-1996.

PF 15-MAR-1996; AU0149.

PR 16-MAR-1995; AU-001789.

PA (UYME) UNIV MELBOURNE.

PI Doughty SW, Lee R, Walker J;

DR WPI; 96-433763/43.

PT Putative protective antigens against Mycoplasma - used for the

detection, prevention or treatment of Mycoplasma infections, esp. M.

PS Claim 13; Page 28; 43pp; English.

CC A 46-48 kDa putative protective antigen against Mycoplasma contains

the N-terminal sequence given in W01034-36. The antigen was isolated from

fragments given in W01034-36. The antigen was isolated from

Mycoplasma hyopneumoniae cells using antibody probes enriched with

FT /note= "codon 254 in the nucleotide sequence is

is a stop codon"

PN W09628472-A1.

PD 19-SEP-1996.

PF 15-MAR-1996; AU0149.

PR 16-MAR-1995; AU-001789.

PA (UYME) UNIV MELBOURNE.

PI Doughty SW, Lee R, Walker J;

DR WPI; 96-433763/43.

DR N-PSDB; T38241.

PT Putative protective antigens against Mycoplasma - used for the

detection, prevention or treatment of Mycoplasma infections, esp. M.

PS Disclosure; Fig 7; 43pp; English.

CC The gene (T38241) coding for a 48 kDa putative protective antigen

(W01037) was obtd. from a Mycoplasma hyopneumoniae genomic library

by screening with a probe generated by PCR amplification (see also

T3814-16). The antigen had originally been isolated from M.

CC hyopneumoniae cells using antibody probes enriched with Mycoplasma-

specific antibodies. Other protective antigens were also identified

(see also W01024-31). Protective antigens and antibodies can be

used in vaccines for preventing or treating mycoplasma infections,

CC partic. M. hyopneumoniae infections in swine. They can also be used

for diagnosis.

CC Sequence 419 AA;

Query Match 100.0%; Score 78; DB 1; Length 419;

Best Local Similarity 100.0%; Pred. No. 1.7e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEQAITKLEGFDTQ 16

DB 301 AEQAITKLEGFDTQ 316

RESULT 5

R70419

ID R70419 standard; Protein; 1456 AA.

AC R70419;

DT 14-FEB-1996 (first entry)

DE Rat homologue of human Wilson disease gene ATP7B.

KW Copper transport; ATPase; Wilson disease; chromosomal markers; probe;

yeast artificial chromosome library; hybridisation; Menkes disease;

KW copper binding site; phosphate domain; transduction; transmembrane;

KW phosphorylation; ATP binding domain; autosomal recessive disorder;

KW toxicity; liver; brain; gene therapy; heavy metal; rat.

OS Rattus rattus.

FT Key

Location/Qualifiers

domain 68..97

FT /label= copper binding domain 1

FT domain 153..182

FT /label= copper binding domain 2

FT domain 267..296

FT /label= copper binding domain 3

FT domain 492..521

FT /label= copper binding domain 5

FT domain 568..597

FT /label= copper binding domain 6

FT domain 648..671

FT /label= transmembrane domain 1

FT domain 683..711

FT /label= transmembrane domain 2

FT domain 726..752

FT /label= transmembrane domain 3

FT domain 761..780

FT /label= transmembrane domain 4

FT domain 834..862

FT /label= energy transduction domain

FT domain 919..947

FT /label= transmembrane domain 5

FT domain 962..1011

FT /label= ion channel/transmembrane domain 6

FT	domain	1012..1033
FT	/label= phosphorylation domain	
FT	domain	1034..1067
FT	/label= transmembrane domain 7	
FT	domain	1170..1190
FT	/label= transmembrane domain 8	
FT	domain	1233..1284
FT	/label= ATP binding domain	
FT	/note= "designated ATP-hinge in specification"	
FT	domain	1315..1362
FT	/label= transmembrane domain 9	
PN	W09508641-A.	
PD	30-MAR-1995.	
PF	21-SEP-1994; CA0519.	
PR	21-SEP-1993; CA106602.	
PR	21-OCT-1993; CA-108927.	
PA	(HSCR-) HSC RES & DEV LP.	
PI	Bull P, Cox DW, Thomas G;	
PI	WPI; 95-139609/18.	
DR	N-PSDB; Q85956.	
DR	Nucleotide sequence contg. the gene for Wilson disease - used to treat and detect Wilson disease	
PT	Example 5; Fig 17; 127pp; English.	
PS	The amino acid sequence of the protein encoded by the rat gene which is homologous to the human gene (Q86541) encoding the copper transporting ATPase (ATP7B) associated with Wilson disease (WD). The rat gene was isolated from a rat liver cDNA library with probes derived from the human WD gene sequence. Twenty-five clones were isolated which resulted in a consensus sequence of about 4.7 kb. The rat gene encodes a single large open reading frame (ORF) with about 300 bp of 3' untranslated sequence. The rat and human gene share about 82% sequence homology at the amino acid level. The overall structure of the gene is similar to the human gene with the exception of the lack of copper binding domain 4 in the rat sequence. The rat gene disorder is found in Long-Evans Cinnamon rats which share many of the clinical and biochemical features with WD, so can be used as models for WD. The gene can be used to treat Wilson disease, an autosomal recessive disorder of copper transport which results in copper accumulation and toxicity to the liver and brain. The DNA sequence can also be used in gene therapy to remove heavy metals from an organ.	
CC	Sequence	1456 AA;
SQ		
Query Match	52.6%; Score 41; DB 1; Length 1456;	
Best Local Similarity	43.8%; Pred. No. 39;	
Matches	7; Conservative	5; Mismatches
Indels	4;	Gaps
Qy	1 AEQAITKLEGGDTQ 16	
Db	144 AQEAVKLVEGTCQ 159	
RESULT	6	
R72343		
ID	R72343 standard; Protein; 1411 AA.	
AC	R72343;	
DE	23-NOV-1995 (first entry)	
DT	Wilson disease protein ATP7B	
KW	Copper transport; ATPase; Wilson disease; chromosomal markers; probe; yeast artificial chromosome library; hybridisation; Menkes disease;	
KW	copper binding site; phosphate domain; transduction; transmembrane; phosphorylation; ATP binding domain; autosomal recessive disorder;	
KW	toxicity; liver; brain; gene therapy; heavy metal.	
OS	Homo sapiens.	
FH	Key Location/Qualifiers	
FT	domain	32..61
FT	/label= copper binding domain 1	
FT	domain	117..146
FT	/label= copper binding domain 2	
FT	domain	231..260
FT	/label= copper binding domain 3	
FT	domain	334..363
FT	/label= copper binding domain 4	
FT		
misc_difference 318	/note= "encoded by CAC"	
domain	/label= copper binding domain 5	
domain	539..568	
/label= copper binding domain 6		
misc_difference 621	/note= "encoded by AAG"	
misc_difference 622	/note= "encoded by TCT"	
misc_difference 770	/note= "encoded by ACA"	
domain	806..833	
/label= energy transduction domain		
domain	940..982	
/label= transmembrane domain		
domain	983..1004	
/label= phosphorylation domain		
domain	1209..1260	
/label= ATP binding domain		
/note= "designated ATP-hinge in specification"		
W09508641-A.		
30-MAR-1995.		
21-SEP-1994; CA0519.		
21-SEP-1993; CA106602.		
21-OCT-1993; CA-108927.		
(HSCR-) HSC RES & DEV LP.		
Bull P, Cox DW, Thomas G;		
WPI; 95-139609/18.		
N-PSDB; Q86541.		
Nucleotide sequence contg. the gene for Wilson disease - used to treat and detect Wilson disease		
Example 2; Fig 10; 127pp; English.		
The amino acid sequence of the copper transporting ATPase (ATP7B) encoded by the Wilson disease (WD) gene. The gene has been located to chromosomal band q14.3, especially between the markers D1S31 and D1S359. The three novel markers D1S314, D1S313 and D1S316 have been located to within 300 kb of the WD gene. The WD gene was isolated from a yeast artificial chromosome (YAC) library by low stringency hybridisation with a probe derived by amplifying a region of the related copper transport gene involved in Menkes disease. The protein encoded by the WD gene contains 6 copper binding sites, a phosphatase domain, a transduction, several potential transmembrane, phosphorylation and ATP binding domains. The WD gene encodes a transcript of 7.5 kb of which 4.2 kb encodes the protein. The transcript is divided into 21 exons. The ATP7B coding regions spans around 40 kb of genomic DNA. The gene can be used to treat Wilson disease, an autosomal recessive disorder of copper transport which results in copper accumulation and toxicity to the liver and brain. The DNA sequence can also be used in gene therapy to remove heavy metals from an organ.		
Sequence	1411 AA;	
SQ		
Query Match	52.6%; Score 41; DB 1; Length 1411;	
Best Local Similarity	43.8%; Pred. No. 37;	
Matches	7; Conservative	5; Mismatches
Indels	4;	Gaps
Qy	1 AEQAITKLEGGDTQ 16	
Db	108 AQEAVKLVEGTCQ 123	
RESULT	7	
R44431		
ID	R44431 standard; Protein; 3567 AA.	
AC	R44431;	
DE	22-DEC-1993 (first entry)	
DT	eryA region polypeptide module #2.	
KW	Saccaropolyspora erythraea; eryA; biosynthesis; polyketide; module;	
KW	erythromycin; condensation; elongation; acyl chain growth;	
KW	gene replacement.	
OS	Saccaropolyspora erythraea.	
PN	W09313663-A.	

PD 22-JUL-1993.
PF 17-JAN-1992; U00427.
PR 17-JAN-1992; WO-U00427.
PA (ABBO) ABBOTT LAB.
PI Donadio S, Katz L, McAlpine JB;
DR WPI; 93-242804/30.
DR N-PSDB; O46806.
PT Biosynthesis of specific polyketide analogues esp. erythromycin
cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT microorganisms
PS Disclosure; Fig 2; 133pp; English.
CC The sequences given in R4430-32 are encoded by the eryA fragment of
the Saccharopolyspora erythraea genome. These polypeptides are
involved in the biosynthesis of the polyketide segment of erythromycin.
eryA is organised in modules and each module takes care of one
condensation step. The precise succession of elongation steps is
dictated by the genetic order of the modules. The DNA encoding
these polypeptides may be specifically altered such that novel
polyketide molecules of desired structure are produced. Three types
of alteration may be produced; those inactivating a single function in
a module which does not arrest acyl chain growth; those inactivating a
single function in a module which does affect chain growth; and those
affecting an entire module. The mutations may be introduced by gene
replacement.
SQ Sequence 3567 AA;

Query Match 48.7%; Score 38; DB 1; Length 3567;
Best Local Similarity 57.1%; Pred. No. 3.9e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAIVTKLEGFDT 15
|||: | :|||
Db 2265 EQAVRGIVLEGFDT 2278

RESULT 8
W80609
ID W80609 standard; Protein; 257 AA.
DT 24-DEC-1998 (first entry)
DE S. pneumoniae protein of unknown function.
KW Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
KW virulence; antibody; infection; detection; treatment; hypothetical;
KW cell wall biosynthetic, external target; minimal gene set protein.
OS Streptococcus pneumoniae.
PN W09826072-A1.
PD 18-JUN-1998.
PF 09-DEC-1997; U22578.
PR 13-DEC-1996; US-036281.
PA (ELIL) LILLY & CO ELI.
PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
PI Mills BJ, Norris FH, Peery RB, Rostek PK, Rostek PR,
PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,
PI Young Belldo ML;
DR WPI; 98-348529/30.
DR N-PSDB; V65291.
PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT for evaluating gene expression, and identification of virulence
PT genes
PS Claim 3; Pages 166-167; 333pp; English.
CC This sequence represents a Streptococcus pneumoniae protein of unknown
function. The invention provides DNA sequences (V65201 to V65304) from
the Streptococcus pneumoniae genome and corresponding protein sequences
(W80605 to W80728). The protein sequences are classified as hypothetical,
cell wall biosynthetic, external target, or minimal gene set proteins. A
recombinant host containing a vector comprising any of the above nucleic
acids can be used for the recombinant expression of the proteins. The
invention also provides a DNA chip having arrayed on it at least 15 base
pair fragment of any one or more of these DNA sequences. The DNA chip can
be used methods for evaluating gene expression in S. pneumoniae and for
identifying virulence genes in S. pneumoniae. Antibodies that selectively
bind to the above proteins or peptide fragments can be used to treat

CC S. pneumoniae infection. The antibodies can also be used to detect
CC S. pneumoniae cells.
SQ Sequence 257 AA;

Query Match 48.7%; Score 38; DB 1; Length 257;
Best Local Similarity 42.9%; Pred. No. 18;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEQAIVTKLEGFDT 14
||: | :|||
Db 26 AEMGVQTMNIVGFD 39

RESULT 9
R71345
ID R71345 standard; Peptide; 72 AA.
AC R71345;
DT 04-OCT-1995 (first entry)
DE Carboxy terminal boundaries of metal-binding site of Wilson's disease
DE protein region WD-2
KW Wilson's disease; chromosome 13; metal binding site.
OS Homo sapiens.
PN W09506714-A.
PD 09-MAR-1995.
PF 01-SEP-1994; U09851.
PR 01-SEP-1993; US-118441.
PA (GEHO) GEN HOSPITAL CORP.
PA (UYCO) UNIV COLUMBIA NEW YORK.
PI Gilliam TC, Tanzi RE;
DR WPI; 95-115430/15.
PT Isolated Wilson's disease nucleic acid mol. - also probes,
PT vectors, etc., useful for diagnosis and gene therapy of Wilson's
PT disease.
PS Example; Figure 15C; 175pp; English.
CC In the N-terminal segment of the Wilson's disease gene product
there are six copper binding motifs, GMXCXC, which are repeated
with 70-80 AA residues spacings. Carboxy terminal boundaries of
metal binding sites were determined by pair by pair alignment of
WD protein copper-binding repeat. Two representative alignments
are shown in Figure 15C together with the consensus sequence (see
R71345, R71346 and R71347.
SQ Sequence 72 AA;

Query Match 47.4%; Score 37; DB 1; Length 72;
Best Local Similarity 40.0%; Pred. No. 6;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAIVTKLEGFDTQ 16
::|: | :|||
Db 1 QEAVVKLVGEGMTCQ 15

RESULT 10
R88320
ID R88320 standard; Protein; 847 AA.
AC R88320;
DT 29-MAR-1996 (first entry)
DE IL-4 Stat peptide.
KW Interleukin-4 signal transducer and activator of transcription;
KW IL-4 Stat; transcription factor; immunosuppressive.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..40
FT /note= "amino acids 1-40 constitute an active
IL-4 Stat peptide (Claim 2, page 21)."
FT peptide 401..650
FT /note= "amino acids 401-650 constitute an active
IL-4 Stat peptide (Claim 3, page 21)."
FT EP-692488-A2.
PN 17-JAN-1996.
PF 05-JUL-1995; 304715.

```

PR 05-JUL-1994; US-259604.
PR 15-JUL-1994; US-276099.
PA (TULA-) TULARIK INC.
PI HOU J, McKnight SL;
DR WPI: 96-070143/08.
DR N-PSDB: T03679.
PT IL-4 signal transducer and activator of transcription (IL-4 Stat)
PT peptides) - bind to natural intracellular IL-4 Stat binding target
PT and are useful to identify cpds. for treatment and diagnosis of
PT immune diseases
PS Disclosure: Page 17-19; 22pp; English.
CC Interleukin-4 signal transducer and activator of transcription (IL-4
CC Stat) (R88320) is a novel transcription factor characterized by
CC selective binding to intracellular domains of cytokine receptors.
CC It was obt'd. by expression of a cDNA clone (T03679) isolated from
CC human monocyte Thp-1 cells. The IL-4 Stat peptide is used to
CC identify cpds. useful in the treatment/diagnosis of immune diseases.
CC bacterial, viral or fungal infections, metabolic or genetic diseases,
CC neoplasia, inflammation etc.
SQ Sequence 847 AA;

Query Match 47.4%; Score 37; DB 1; Length 847;
Best Local Similarity 72.7%; Pred. No. 1.1e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EQAITKLEGG 12
   II : III III
Db 799 EQDLTKLLEGG 809

RESULT 11
W21894
ID W21894 standard; Protein; 432 AA.
AC W21894;
DT 11-SEP-1997 (first entry)
DE Asparaginyl-tRNA synthetase from Staphylococcus aureus.
KW tRNA synthetase; Bacillus subtilis; immunological response;
KW antibody; bacterial infection; adherence; damaged tissue; wound healing;
KW skin; protection; vaccine.
OS Staphylococcus aureus.
PN EP-785259-A1.
PD 23-JUL-1997; 300308.
PF 17-JAN-1997; 300308.
PR 19-JAN-1996; GB-001099.
PR 19-JAN-1996; GB-001096.
PR 27-JUL-1996; GB-015845.
PR 30-OCT-1996; GB-022617.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
PI Hodgson JE, Lawlor EJ;
DR N-PSDB; T73692.
PT DNA encoding asparaginyl-tRNA synthetase from Staphylococcus aureus
PT WCUH 29 - useful for protection against bacterial infections
PS Claim 1; Pages 19-21; 33pp; English.
CC The present sequence is a novel asparaginyl-tRNA synthetase protein
CC from Staphylococcus aureus WCUH29 (NCIMB 40771), which is related by
CC amino acid sequence homology to Bacillus subtilis asparaginyl-tRNA
CC synthetase. The enzyme catalyses the aminoacylation of tRNA-Asn, by
CC a two step mechanism. The first step involves formation of a stable
CC enzyme-asparaginyl adenylate complex resulting from the specific binding
CC and reaction of ATP and L-asparagine. Subsequently, the 3' terminal
CC adenosine of enzyme-bound tRNA-Asn reacts with the aminoacyl adenylate,
CC leading to the esterification of the tRNA and release of AMP. Vectors
CC comprising the DNA (or polynucleotides having at least 70 % identity to
CC it) can be used for the recombinant production of the enzyme. The enzyme
CC or its related DNA (through gene therapy) is used to induce an
CC immunological response in a mammal to generate antibodies to protect
CC against disease. The antibodies protect against invasion of bacteria,
CC e.g. by blocking adherence of bacteria to damaged tissue, including
CC wounds in skin or connective tissue caused by mechanical, chemical or
CC thermal damage or by implantation of in-dwelling devices, or wounds in
CC the mucous membranes.

SQ Sequence 432 AA;

Query Match 47.4%; Score 37; DB 1; Length 432;
Best Local Similarity 61.5%; Pred. No. 50;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITKLEGGFD 14
   : II : III III
Db 282 DDAIEFLKAEFGD 294

RESULT 12
W63842
ID W63842 standard; Protein; 432 AA.
AC W63842;
DT 01-OCT-1998 (first entry)
DE S. aureus asparaginyl tRNA synthetase protein #1.
KW Asparaginyl tRNA synthetase; catalytic; protein synthesis; disease;
KW invasive infection; bacteraemia; osteomyelitis; septic; arthritis;
KW thrombophlebitis; acute bacterial endocarditis; screening; antibiotic;
KW vaccine.
OS Staphylococcus aureus.
PN US5789217-A.
PD 04-AUG-1998.
PF 17-JAN-1997; 785076.
PR 17-JAN-1997; US-785076.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
PI Hodgson JE, Lawlor EJ;
DR WPI: 98-446083/38.
DR N-PSDB; V44040.
PT DNA encoding asparaginyl tRNA synthetase from Staphylococcus aureus
PT - useful for recombinant production of the enzyme in expression
PT systems, to screen for antibiotics
PS Claim 1a; Fig 2; 22pp; English.
CC This sequence represents a novel Staphylococcus aureus asparaginyl tRNA
CC synthetase. This protein can be used to produce recombinant proteins
CC or fragments of it that retain catalytic activity. The enzyme is used in
CC S. aureus during protein synthesis. S. aureus is responsible for a variety
CC of disease conditions in patients e.g. invasive infections, bacteraemia
CC in cancer patients, osteomyelitis, septic arthritis, septic
CC thrombophlebitis and acute bacterial endocarditis. The polynucleotides
CC and expression systems that produce them can also therefore be used to
CC screen for antibiotics against S. aureus. The polypeptides can also be
CC used to raise antibodies, either for use as a vaccine or in immunogenic
CC detection methods.
SQ Sequence 432 AA;

Query Match 47.4%; Score 37; DB 1; Length 432;
Best Local Similarity 61.5%; Pred. No. 50;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITKLEGGFD 14
   : II : III III
Db 282 DDAIEFLKAEFGD 294

RESULT 13
W63843
ID W63843 standard; Protein; 430 AA.
AC W63843;
DT 01-OCT-1998 (first entry)
DE S. aureus asparaginyl tRNA synthetase protein #2.
KW Asparaginyl tRNA synthetase; catalytic; protein synthesis; disease;
KW invasive infection; bacteraemia; osteomyelitis; septic; arthritis;
KW thrombophlebitis; acute bacterial endocarditis; screening; antibiotic;
KW vaccine.
OS Staphylococcus aureus.
PN US5789217-A.
PD 04-AUG-1998.
PF 17-JAN-1997; 785076.
PR 17-JAN-1997; US-785076.

```


Search completed: November 13, 1999, 18:59:48
Job time: 127 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:27 ; Search time 64.87 Seconds
(without alignments)
2.819 Million cell updates/sec

Title: US-08-913-430-5

Perfect score: 78

Sequence: 1 AEOATKLEGFDTQ 16

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/PCTUS9_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	48.7	73	1	US-08-280-443-25
2	38	48.7	73	1	US-08-457-459-25
3	38	48.7	73	2	US-08-555-678-25
4	38	48.7	3567	2	US-07-642-734C-4
5	38	48.7	73	3	PCT-US95-02275-25
6	37	47.4	847	1	US-08-276-099A-2
7	37	47.4	847	1	US-08-781-890-2
8	37	47.4	432	2	US-08-785-076-2
9	37	47.4	430	2	US-08-785-076-3
10	36	46.2	2101	1	US-08-466-390-4
11	36	46.2	2101	1	US-08-470-950-4
12	36	46.2	2101	2	US-08-467-781-4
13	36	46.2	2101	2	US-08-195-487-4
14	36	46.2	2101	2	US-08-483-924-4
15	36	46.2	2101	3	PCT-US93-06160-4
16	35	44.9	323	1	US-07-992-827D-1
17	35	44.9	323	1	US-08-216-593-1
18	35	44.9	189	2	US-08-064-694-2
19	35	44.9	981	2	US-08-649-046-2
20	35	44.9	323	3	PCT-US93-12380-1
21	34	43.6	526	1	US-08-298-428-4
22	33	42.3	1708	1	US-08-493-092-2
23	33	42.3	1708	2	US-08-508-836A-2
24	33	42.3	3056	2	US-08-508-836A-8
25	33	42.3	3056	2	US-08-629-001A-3
26	33	42.3	3056	2	US-08-874-266-2
27	32.5	41.7	876	2	US-08-785-429-2
28	32	41.0	424	1	US-08-045-269C-2
29	32	41.0	1445	1	US-08-015-986A-2
30	32	41.0	1442	1	US-08-015-986A-3
31	32	41.0	267	1	US-08-015-973-3
32	32	41.0	406	1	US-07-973-431B-1
33	32	41.0	1132	1	US-08-446-038B-18
34	32	41.0	269	1	US-08-241-766-6
35	32	41.0	269	1	US-08-241-766-7
36	32	41.0	1154	1	US-08-357-598-7
37	32	41.0	1132	1	US-08-446-010B-18
38	32	41.0	1154	1	US-08-446-010B-24
39	32	41.0	1142	1	US-08-097-997A-11

40 32 41.0 1153 1 US-08-097-997A-14 Sequence 14, Appl
41 32 41.0 915 1 US-08-453-862-2 Sequence 2, Appl
42 32 41.0 1132 2 US-08-805-445-18 Sequence 18, Appl
43 32 41.0 472 2 US-08-272-255-10 Sequence 10, Appl
44 32 41.0 15281 2 US-08-471-119A-2 Sequence 2, Appl
45 32 41.0 915 2 US-08-452-734A-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-280-443-25
; Sequence 25, Application US/08280443
; Patent No. 5643778
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,443
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49AUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 73 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-280-443-25

Query Match 48.7%; Score 38; DB 1; Length 73;
Best Local Similarity 72.7%; Pred. No. 2.3;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEOATKLE 11
|||: |||:
DB 63 AEOALKLE 73

RESULT 2
US-08-457-459-25
; Sequence 25, Application US/08457459
; Patent No. 5677428
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Howson and Howson
;; STREET: Spring House Corporate Cntr, P.O. Box 457
;; CITY: Spring House
;; STATE: Pennsylvania
;; COUNTRY: USA
;; ZIP: 19477
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/457,459
;; FILING DATE:
;; CLASSIFICATION: 435
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/197,794
;; FILING DATE: 17-FEB-1994
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/280,443
;; FILING DATE: 25-JUL-1994
;;
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bak, Mary E.
;; REGISTRATION NUMBER: 31,215
;; REFERENCE/DOCKET NUMBER: WST49CUSA
;; TELEPHONE: 215-540-9206
;; TELEFAX: 215-540-5818
;;
;; INFORMATION FOR SEQ ID NO: 25:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 73 amino acids
;; TYPE: amino acid
;; TOPOLOGY: unknown
;; MOLECULE TYPE: protein
;;
;; US-08-457-459-25

Query Match 48.7%; Score 38; DB 1; Length 73;
Best Local Similarity 72.7%; Pred. No. 2.3;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEQAATKLE 11
| | | | | | | | | |
Db 63 AEQAATKLE 73

RESULT 3
US-08-555-678-25
; Sequence 25, Application US/08555678
; Patent No. 5763174
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods
; TITLE OF INVENTION: of use Thereof
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,678
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/197,794
;; FILING DATE: 17-FEB-1994
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/280,443
;; FILING DATE: 25-JUL-1994
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/457,459
;; FILING DATE: 01-JUN-1995
;;
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bak, Mary E.
;; REGISTRATION NUMBER: 31,215
;; REFERENCE/DOCKET NUMBER: WST49DUSA
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-540-9206
;; TELEFAX: 215-540-5818
;;
;; INFORMATION FOR SEQ ID NO: 25:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 73 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: unknown
;; MOLECULE TYPE: protein
;;
;; US-08-555-678-25

Query Match 48.7%; Score 38; DB 2; Length 73;
Best Local Similarity 72.7%; Pred. No. 2.3;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEQAATKLE 11
| | | | | | | | | |
Db 63 AEQAATKLE 73

RESULT 4
US-07-642-734C-4
; Sequence 4, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Danckers, Andreas M
; REGISTRATION NUMBER: 32652
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-938-2623
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: amino acid

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-642-734C-4

Query Match 48.7%; Score 38; DB 2; Length 3567;
Best Local Similarity 57.1%; Pred. No. 2.1e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITKLEGGDT 15
   ||| | : |||
Db 2265 EQAVRGLVEGGDT 2278

RESULT 5
PCT-US95-02275-25
; Sequence 25, Application PC/TUS9502275
; GENERAL INFORMATION:
; APPLICANT: Wistar Institute of Anatomy & Biology
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02275
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,443
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49BPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 73 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
PCT-US95-02275-25

Query Match 48.7%; Score 38; DB 3; Length 73;
Best Local Similarity 72.7%; Pred. No. 2.3;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEQAITKLE 11
   ||| | : |||
Db 63 AEQAALKLE 73

RESULT 6
US-08-276-099A-2
; Sequence 2, Application US/08276099A
```

```
; Patent No. 5591825
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L
; APPLICANT: Hou, Jinzhao
; TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
; TITLE OF INVENTION: BINDING ASSAYS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,099A
; FILING DATE: 15-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard Aron
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59451-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 847 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-276-099A-2

Query Match 47.4%; Score 37; DB 1; Length 847;
Best Local Similarity 72.7%; Pred. No. 60;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EQAITKLEGG 12
   || : ||| |||
Db 799 EQDLTKLLEGG 809

RESULT 7
US-08-781-890-2
; Sequence 2, Application US/08781890
; Patent No. 5710266
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L
; APPLICANT: Hou, Jinzhao
; TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
; TITLE OF INVENTION: BINDING ASSAYS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,890
```

;; FILING DATE: 05-JAN-1997
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/276,099
;; FILING DATE: 15-JUL-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Osman, Richard Aron
;; REGISTRATION NUMBER: 36,627
;; REFERENCE/DOCKET NUMBER: A-59451-1/RAO
;; TELEPHONE: (415) 781-1989
;; TELEFAX: (415) 398-3249
;; TELEX: 910 277299
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 847 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-781-890-2

Query Match 47.4%; Score 37; DB 1; Length 847;
Best Local Similarity 72.7%; Pred. No. 60;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EQAITKLEKLEG 12
; : : : : :
Db 799 EQDLTKLLEK 809

RESULT 8
US-08-785-076-2
; Sequence 2, Application US/08785076
; Patent No. 5789217
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 5789217el tRNA Synthetase
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,076
; FILING DATE: 17-JAN-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9601096.2
; FILING DATE: 19-JAN-1996
; APPLICATION NUMBER: 9615845.6
; FILING DATE: 27-JUL-1996
; APPLICATION NUMBER: 9622617.0
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31354-3
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 847 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-781-890-2

;; LENGTH: 432 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-785-076-2

Query Match 47.4%; Score 37; DB 2; Length 432;
Best Local Similarity 61.5%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITKLEKLEG 14
; : : : : :
Db 282 DDAIEFLKAEKGF 294

RESULT 9
US-08-785-076-3
; Sequence 3, Application US/08785076
; Patent No. 5789217
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 5789217el tRNA Synthetase
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,076
; FILING DATE: 17-JAN-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9601096.2
; FILING DATE: 19-JAN-1996
; APPLICATION NUMBER: 9615845.6
; FILING DATE: 27-JUL-1996
; APPLICATION NUMBER: 9622617.0
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31354-3
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-785-076-3

Query Match 47.4%; Score 37; DB 2; Length 430;
Best Local Similarity 61.5%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITKLEKLEG 14
; : : : : :
Db 282 DDAIEFLKAEKGF 294

```

Db      280 DDIAIEFLKAEFGD 292

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/470,950
;   FILING DATE: 06-JUN-1995
;   CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
;   NAME: PITCHER ESQ, EDMUND R
;   REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 248-7000
;   TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2101 amino acids
;   TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-950-4

Query Match          46.2%; Score 36; DB 1; Length 2101;
Best Local Similarity 53.3%; Pred. No. 2.6e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0

Qy      2 EQAITKLKGEGDTQ 16
       |||::|||
Db     1340 EQALSTLQLEHTSQ 1354

RESULT 12
US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOKUTATY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/467,781
;   FILING DATE: 06-JUN-1995
;   CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
;   NAME: PITCHER ESQ, EDMUND R
;   REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 248-7000
;   TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2101 amino acids
;   TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-781-4

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/466,390
;   FILING DATE: 06-JUN-1995
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: PITCHER ESQ, EDMUND R
;   REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 248-7000
;   TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2101 amino acids
;   TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-390-4

Query Match          46.2%; Score 36; DB 1; Length 2101;
Best Local Similarity 53.3%; Pred. No. 2.6e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      2 EQAITKLKGEGDTQ 16
       |||::|||
Db     1340 EQALSTLQLEHTSQ 1354

RESULT 11
US-08-470-950-4
; Sequence 4, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOKUTATY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

[illegible]

Query Match 46.2%; Score 36; DB 2; Length 2101;
Best Local Similarity 53.3%; Pred. No. 2.6e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITKLEGFDTQ 16
| | | | : | | | | |
Db 1340 EQALSTLQLEHTSTQ 1354

RESULT 13
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESO, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELEPHONE: 617/248-7100
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-487-4

Query Match 46.2%; Score 36; DB 2; Length 2101;
Best Local Similarity 53.3%; Pred. No. 2.6e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITKLEGFDTQ 16
| | | | : | | | | |
Db 1340 EQALSTLQLEHTSTQ 1354

RESULT 14
US-08-483-924-4
; Sequence 4, Application US/08483924
; Patent No. 5862876
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; INTERIOR NUCLEAR MATRIX

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESO, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-924-4

Query Match 46.2%; Score 36; DB 2; Length 2101;
Best Local Similarity 53.3%; Pred. No. 2.6e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITKLEGFDTQ 16
| | | | : | | | | |
Db 1340 EQALSTLQLEHTSTQ 1354

RESULT 15
PCT-US93-06160-4
; Sequence 4, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06160
; FILING DATE: 19930621
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESO, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100

```
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-06160-4
```

```
Query Match 46.2%; Score 36; DB 3; Length 2101;
Best Local Similarity 53.3%; Pred. No. 2.6e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITYKLEGFDTQ 16
    |||:: |||
Db 1340 EQALSTLQLEHTSTQ 1354
```

Search completed: November 13, 1999, 10:56:29
Job time: 1358 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:33 ; Search time 251.81 Seconds
(without alignments)
4.022 Million cell updates/sec

Title: US-08-913-430-5

Perfect score: 78

Sequence: 1 AEQATKLEGFDTQ 16

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA:*
1: /cgn2_6/ptodata/1/paa/PCTUS99_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US094A_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	100.0	419	4	US-08-183-774-1
2	78	100.0	419	13	US-08-913-430-2
3	78	100.0	16	13	US-08-913-430-5
4	41	52.6	1485	10	US-08-617-833-2
5	41	52.6	1426	10	US-08-617-833D-2
6	41	52.6	283	15	US-09-107-532-6895
7	40.5	51.9	295	15	US-09-134-000-4352
8	40	51.3	268	16	US-09-248-796-17058
9	40	51.3	268	19	US-60-096-409-17058
10	39	50.0	397	24	US-60-150-582-507
11	38	48.7	257	1	PCT-US97-22578-96
12	38	48.7	3567	2	US-07-642-734B-4
13	38	48.7	73	4	US-08-197-794-25
14	38	48.7	73	4	US-08-197-794A-25
15	38	48.7	3567	7	US-08-439-009-4
16	38	48.7	3567	7	US-08-439-009A-4
17	38	48.7	185	10	US-08-625-811-1573
18	38	48.7	257	13	US-08-987-147-2
19	38	48.7	249	15	US-09-107-532-4701

20	37	47.4	311	12	US-08-827-356-4962
21	37	47.4	432	13	US-08-913-579-2
22	37	47.4	430	13	US-08-913-579-3
23	37	47.4	432	14	US-09-018-824-2
24	37	47.4	430	14	US-09-018-824-3
25	37	47.4	847	14	US-09-087-465-12
26	37	47.4	169	15	US-09-134-000-5112
27	37	47.4	1196	16	US-09-248-796-19876
28	37	47.4	234	16	US-09-270-767-42924
29	37	47.4	1196	19	US-60-096-409-19876
30	36	46.2	933	1	PCT-US99-07643-10
31	36	46.2	933	1	PCT-US99-10843-22
32	36	46.2	1152	1	PCT-US99-18298-46
33	36	46.2	2272	8	US-08-478-408-2
34	36	46.2	2272	8	US-08-478-408-4
35	36	46.2	2272	8	US-08-484-680-2
36	36	46.2	2272	8	US-08-484-680-4
37	36	46.2	2272	8	US-08-487-289-2
38	36	46.2	2272	8	US-08-487-289-4
39	36	46.2	2272	10	US-08-660-396-2
40	36	46.2	2192	10	US-08-660-396-4
41	36	46.2	359	15	US-09-134-000-4362
42	36	46.2	60	16	US-09-270-767-32933
43	36	46.2	293	16	US-09-270-767-33186
44	36	46.2	159	16	US-09-270-767-33607
45	36	46.2	640	16	US-09-270-767-44307

ALIGNMENTS

RESULT 1

US-08-183-774-1
Sequence 1, Application US/08183774
GENERAL INFORMATION:

APPLICANT: SETO, Yasuhiro

APPLICANT: FUTU, Satoshi

APPLICANT: MITSUSE, Shizuo

APPLICANT: MATSUO, Kanako

APPLICANT: TSUNA, Mika

TITLE OF INVENTION: DNA's Encoding Surface Antigen of

TITLE OF INVENTION: Mycoplasma hyopneumoniae, DNA Fragments for Primer, of

TITLE OF INVENTION: Recombinant Antigenic Peptides and Diagnostic Method of

TITLE OF INVENTION: Mycoplasma Pneumoniae of Swine Using Same

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC Compatible

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/183,774

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/747,015

FILING DATE: 19-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, Norman F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 4183-001-0

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 419 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-183-774-1

Query Match 100.0%; Score 78; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 7.6e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEQAITKLKLEGFDTQ 16
|||||
DB 301 AEQAITKLKLEGFDTQ 316

RESULT 2
US-08-913-430-2
; Sequence 2, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 2
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-2

Query Match 100.0%; Score 78; DB 13; Length 419;
Best Local Similarity 100.0%; Pred. No. 7.6e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEQAITKLKLEGFDTQ 16
|||||
DB 301 AEQAITKLKLEGFDTQ 316

RESULT 3
US-08-913-430-5
; Sequence 5, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 5
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-5

Query Match 100.0%; Score 78; DB 13; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEQAITKLKLEGFDTQ 16
|||||
DB 1 AEQAITKLKLEGFDTQ 16

RESULT 4
US-08-617-833-2
; Sequence 2, Application US/08617833
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: WILSON DISEASE GENE
; NUMBER OF SEQUENCES: 31
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,833
; FILING DATE:
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1485 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-617-833-2

Query Match 52.6%; Score 41; DB 10; Length 1485;
Best Local Similarity 43.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEQAITKLKLEGFDTQ 16
|:::|
DB 167 AOEAVVKLKRVGTCQ 182

RESULT 5
US-08-617-833D-2
; Sequence 2, Application US/08617833D
; GENERAL INFORMATION:
; APPLICANT: COX, DIANE W
; APPLICANT: BULL, PETER
; APPLICANT: THOMAS, GORDON
; TITLE OF INVENTION: WILSON DISEASE GENE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLSON & HIERL, LTD.
; STREET: TWENTY NORTH WACKER DRIVE
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: UNITED STATES
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,833D
; FILING DATE: 21-SEP-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: ROTHSTEIN, SEYMOUR
; REGISTRATION NUMBER: 19,369

REFERENCE/DOCKET NUMBER: 96,387
TELEPHONE: 312-580-1180
TELEFAX: 312-580-1189
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-617-833D-2

Query Match 52.6%; Score 41; DB 10; Length 1426;
Best Local Similarity 43.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEQAATKLEGGDTQ 16
|::|::|::|::|
DB 123 AQAAYKLRVEGTCQ 138

RESULT 6
US-09-107-532-6895
; Sequence 6895, Application US/09107532
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR
; NUMBER OF SEQUENCES: 7308
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:

SOFTWARE:
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085598
FILING DATE: May 14, 1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6895:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1...263
US-09-107-532-6895

Query Match 52.6%; Score 41; DB 15; Length 263;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 2 EQAATKLEGGDT 15
|::|::|::|::|
DB 32 EMGVTKMAVVGFD 45

RESULT 7
US-09-134-000-4352
; Sequence 4352, Application US/09134000A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS
; TITLE OF INVENTION: FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-005
; CURRENT APPLICATION NUMBER: US/09/134,000A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 6810
; SEQ ID NO 4352
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000-4352

Query Match 51.9%; Score 40.5; DB 15; Length 295;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 9; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 1 AEQAATKLEGGDT 15
|::|::|::|::|
DB 85 AQAAYKLRVEGFD 98

RESULT 8
US-09-248-796-17058
; Sequence 17058, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 17058
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-17058

Query Match 51.3%; Score 40; DB 16; Length 268;
Best Local Similarity 50.0%; Pred. No. 30;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 AEQAATKLEGGDT 14
|::|::|::|::|
DB 246 AEKGIKVRMEGLE 259

RESULT 9
US-60-096-409-17058
; Sequence 17058, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A

; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 17058
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-17058

Query Match 51.3%; Score 40; DB 19; Length 268;
Best Local Similarity 50.08; Pred. No. 30;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 AEAQITKLEKLEGF 14
||:|:|:|:|:
DB 246 AEKGIKVKMEGLE 259

RESULT 10

US-60-150-582-507
; Sequence 507, Application US/60150582
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien
; TITLE OF INVENTION: ISOLATED KINASE PROTEINS, NUCLEIC ACID
; FILE REFERENCE: CLO00084
; CURRENT APPLICATION NUMBER: US/60/150,582
; CURRENT FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 688
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 507
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(397)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-60-150-582-507

Query Match 50.0%; Score 39; DB 24; Length 397;
Best Local Similarity 81.8%; Pred. No. 72;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EQAITKLEKLEK 12
|||||
DB 135 EQAIEGLKLEK 145

RESULT 11

PCT-US97-22578-96
; Sequence 96, Application PC/TUS9722578
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Burgett, Stanley G.
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Jaskunas Jr., Stanley R.
; APPLICANT: Mills, Bradley J.
; APPLICANT: Norris, Franklin H.
; APPLICANT: Peery, Robert B.
; APPLICANT: Rostek Jr., Paul R.
; APPLICANT: Skatrud, Paul L.
; APPLICANT: Smith, Michele C.
; APPLICANT: Rocky, Pamela K.
; APPLICANT: Young-Bellido, Michele
; TITLE OF INVENTION: Streptococcus Pneumoniae DNA Sequences
; NUMBER OF SEQUENCES: 228
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana

; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/22578
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11162
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US97-22578-96

Query Match 48.7%; Score 38; DB 1; Length 257;
Best Local Similarity 42.9%; Pred. No. 65;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEAQITKLEKLEGF 14
||:|:|:|:|:
DB 26 AEMGYQMNVGFD 39

RESULT 12

US-07-642-734B-4
; Sequence 4, Application US/07642734B
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; SPECIFIC POLYKETIDES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 6006403500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734B
; FILING DATE: 19910117
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, Daniel W
; REGISTRATION NUMBER: 31912
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-688-7742
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear

MOLECULE TYPE: protein
US-07-642-734B-4

Query Match 48.7%; Score 38; DB 2; Length 3567;
Best Local Similarity 57.1%; Pred. No. 1.6e+03;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EQAITKLEGGFDT 15
| | | | |
Db 2265 EQAVRGVGGFDT 2278

RESULT 13

US-08-197-794-25
Sequence 25, Application US/08197794
GENERAL INFORMATION:
APPLICANT: Nishikura, Kazuko
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/197,794
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST490USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 73 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-197-794-25

Query Match 48.7%; Score 38; DB 4; Length 73;
Best Local Similarity 72.7%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEQAITKLE 11
| | | | |
Db 63 AEQALKLE 73

RESULT 14

US-08-197-794A-25
Sequence 25, Application US/08197794A
GENERAL INFORMATION:
APPLICANT: Nishikura, Kazuko
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P.O. Box 457

CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/197,794A
FILING DATE: 17-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST490USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 73 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-197-794A-25

Query Match 48.7%; Score 38; DB 4; Length 73;
Best Local Similarity 72.7%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEQAITKLE 11
| | | | |
Db 63 AEQALKLE 73

RESULT 15

US-08-439-009-4
Sequence 4, Application US/08439009
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/642734
FILING DATE: 17-JAN-91
ATTORNEY/AGENT INFORMATION:
NAME: Danckers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 708-937-9396
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-439-009-4

```

```

Query Match      48.7%  Score 38; DB 7; Length 3567;
Best Local Similarity 57.1%  Pred. NO. 1.6e+03;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 2 EQAITKIKLEGFDT 15
Db 2265 EQAVRGLVEQGFDT 2278

```

```

Search completed: November 13, 1999, 05:08:34
Job time: 10860 sec

```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:07:55 ; Search time 75.45 Seconds
(without alignments)
8.496 Million cell updates/sec

Title: US-08-913-430-5
Perfect score: 78
Sequence: 1 AEQAITKLEGFDTQ 16

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	78	100.0	419	2	A56153	46K surface antige
2	43	55.1	550	2	A71668	60 kd chaperonin (
3	41	52.6	1465	1	S78555	copper-transprotin
4	41	52.6	1411	2	S40525	copper-transprotin
5	41	52.6	1451	2	I58124	copper-transprotin
6	41	52.6	501	2	A42609	lysine--trNA ligas
7	41	52.6	92	2	JX0089	subtilisin inhibit
8	41	52.6	483	2	A71067	hypothetical prote
9	40	51.3	294	2	E70629	formate dehydrogen
10	39	50.0	916	2	E71330	probable preprotel
11	39	50.0	441	2	C64821	conserved hypothet
12	39	50.0	1172	2	S51623	cut14 protein - fi
13	39	50.0	275	2	C70814	hypothetical prote
14	39	50.0	575	2	C34106	protein kinase (EC
15	39	50.0	614	2	D34106	protein kinase (EC
16	38.5	49.4	119	2	S77769	methylenetetrahydr
17	38	48.7	226	1	NREB3	ribonuclease III (
18	38	48.7	259	1	TPBPBR	terminal protein -
19	38	48.7	541	2	JX0112	signal recognition
20	38	48.7	188	2	H71875	hypothetical prote
21	38	48.7	296	2	E64789	ybcH protein - Esc
22	38	48.7	440	2	D71715	hypothetical prote
23	38	48.7	257	2	D70079	gluconate 5-dehydr
24	38	48.7	3573	2	S23070	erythronolide synt
25	38	48.7	222	2	F69335	conserved hypothet
26	38	48.7	607	2	C69503	conserved hypothet
27	37.5	48.1	319	2	S40876	cell division prot
28	37	47.4	496	2	S68160	probable RNA bindi
29	37	47.4	482	2	S77660	ribosomal protein
30	37	47.4	227	2	C64159	transcription fact
31	37	47.4	150	2	S65996	transcription regu
32	37	47.4	583	2	T01470	pyrophosphate--fru
33	37	47.4	109	2	S19401	hypothetical prote
34	37	47.4	461	2	S54477	hypothetical prote
35	37	47.4	848	2	A54740	interleukin-4-indu
36	36.5	46.8	248	2	F71026	hypothetical prote
37	36	46.2	702	1	SHEGCD	guanosine 3',5'-bi
38	36	46.2	102	1	QPTK	acylphosphatase (E
39	36	46.2	102	1	QPCH	acylphosphatase (E

ALIGNMENTS

RESULT 1

A56153

46K surface antigen precursor - Mycoplasma hyopneumoniae (SGC3)

C:Species: Mycoplasma hyopneumoniae

C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 09-Sep-1997

C:Accession: A56153

R:Puto, S.; Seto, Y.; Mitsuse, S.; Mori, Y.; Suzuki, T.; Kawai, K.

J. Bacteriol. 177, 1915-1917, 1995

A:Title: Molecular cloning of a 46-kilodalton surface antigen (P46) gene from Mycopla

A:Reference number: A56153; MUID:95204368

A:Accession: A56153

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-419 <FUT>

A:Cross-references: GB:D16682; NID:g517134; PID:d1004604; PID:g1220114

C:Genetics:

A:Gene: P46

A:Genetic code: SGC3

C:Keywords: surface antigen

Query Match 100.0%; Score 78; DB 2; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEQAITKLEGFDTQ 16

|||||

Db 301 AEQAITKLEGFDTQ 316

RESULT 2

A71668

60 kd chaperonin (groEL) RP626 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 29-Jan-1999

C:Accession: A71668

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark

Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630

A:Accession: A71668

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-550 <AND>

A:Cross-references: GB:AJ235272; GB:AJ235269; NID:g3861033; PID:e1342911; PID:g386116

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: groEL; RP626

C:Superfamily: chaperonin groEL

Query Match 55.1%; Score 43; DB 2; Length 550;
Best Local Similarity 56.2%; Pred. No. 5.1;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 AEQAITKLEGFDTQ 16

.

Db 422 ASQTLKLVENKDDQ 437

RESULT 3

S78555

S40525

copper-transporting ATPase (EC 3.6.1.-) beta chain - human

N:Alternate names: Cu2+-transporting ATPase P-type; Wilson disease protein

C:Species: Homo sapiens (man)

C:Date: 07-Oct-1994 #sequence_revision 17-Nov-1995 #text_change 15-Jan-1999

C:Accession: S40525

R:Bull, P.C.; Thomas, G.R.; Rommens, J.M.; Forbes, J.R.; Wilson Cox, D.

Nature Genet. 5, 327-337, 1993

A:Title: The Wilson disease gene is a putative copper transporting P-type ATPase similar to the mouse ATP7B

A:Reference number: S40525

A:Accession: S40525

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1411 <BUL>

C:Genetics:

A:Gene: GDB:ATP7B; WND

A:Cross-references: GDB:120494; OMIM:277900

A:Map position: 13q14.3-13q21.1

C:Superfamily: human copper-transporting ATPase; ATPase nucleotide-binding domain homologous to ATPase; transmembrane protein; Wilson disease

C:Keywords: copper transport; hydrolase; ion transport; transmembrane protein; Wilson disease

F:32-61/Domain: heavy-metal-associated homology <HMA1>

F:670-1044/Domain: ATPase transduction domain homology <ATT>

F:1159-1301/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 52.6%; Score 41; DB 2; Length 1411;

Best Local Similarity 43.8%; Pred. No. 32;

Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEQAITYKLEGFDTQ 16

|||||

Db 108 AQEAVVLRVEGTCQ 123

|||||

RESULT 5

I58124

copper-transporting ATPase (EC 3.6.1.-) beta chain - rat

N:Alternate names: Wilson disease protein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 15-Jan-1999

C:Accession: I58124; S46482

R:Wu, J.; Forbes, J.R.; Chen, H.S.; Cox, D.W.

Nature Genet. 7, 541-545, 1994

A:Title: The LEC rat has a deletion in the copper transporting ATPase gene homologous to the mouse ATP7B

A:Reference number: I58124; MUID:95038843

A:Accession: I58124

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1451 <RES>

A:Cross-references: EMBL:U08344; NID:g555675; PID:g555676

R:Yamaguchi, Y.; Helny, M.E.; Shimizu, N.; Aoki, T.; Gitlin, J.D.

Biochem. J. 301, 1-4, 1994

A:Title: Expression of the Wilson disease gene is deficient in the Long-Evans Cinnamon rat

A:Reference number: S46482

A:Accession: S46482

A:Molecule type: mRNA

A:Residues: 530-575, 'AC', 578-584, 'F', 586-605 <VAN>

C:Superfamily: human copper-transporting ATPase; ATPase nucleotide-binding domain homologous to ATPase; transmembrane protein

C:Keywords: copper transport; hydrolase; ion transport; transmembrane protein

F:563-592/Domain: heavy-metal-associated homology <HMA6>

F:694-1068/Domain: ATPase transduction domain homology <ATT>

F:1178-1320/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 52.6%; Score 41; DB 2; Length 1451;

Best Local Similarity 43.8%; Pred. No. 33;

Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEQAITYKLEGFDTQ 16

|||||

RESULT 6

A42609 lysine--tRNA ligase (EC 6.1.1.6) - Campylobacter jejuni
 N:Alternate names: lysyl-tRNA synthetase (LysRS)
 C:Species: Campylobacter jejuni
 C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
 C:Accession: A42609
 R:Chan, V.L.; Bingham, H.L.
 J. Bacteriol. 174, 695-701, 1992
 A:Title: Lysyl-tRNA synthetase gene of Campylobacter jejuni.
 A:Reference number: A42609; MUID:92121106
 A:Accession: A42609
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-501 <CHA>
 A:Cross-references: GB:M63448; NID:g144207; PID:g144208
 A:Note: sequence extracted from NCBI backbone (NCBIN:77414, NCBIP:77415)
 C:Superfamily: lysine-tRNA ligase
 C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 52.6%; Score 41; DB 2; Length 501;
 Best Local Similarity 53.8%; Pred. No. 11;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 EQAITKLKLEGFDTQ 14
 ||| ||| |||
 Db 338 EKILTKLKADGFE 350

RESULT 7

JX0089 subtilisin inhibitor - adzuki bean
 C:Species: Vigna angularis (adzuki bean)
 C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 21-Aug-1998
 C:Accession: JX0089
 R:Nozawa, H.; Yamagata, H.; Aizono, Y.; Yoshikawa, M.; Iwasaki, T.
 J. Biochem. 106, 1003-1008, 1989
 A:Title: The complete amino acid sequence of a subtilisin inhibitor from Adzuki beans (Vigna angularis).
 A:Reference number: JX0089; MUID:90186673
 A:Accession: JX0089
 A:Molecule type: protein
 A:Residues: 1-92 <NOZ>
 C:Superfamily: eglin C

Query Match 52.6%; Score 41; DB 2; Length 92;
 Best Local Similarity 62.5%; Pred. No. 1.7;
 Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AEQAITKLKLEGFDTQ 16
 ||| ||| |||
 Db 41 AEQAETKIKKEEMVDVQ 56

RESULT 8

A71067 hypothetical protein PH1231 - Pyrococcus horikoshii
 C:Species: Pyrococcus horikoshii
 C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
 C:Accession: A71067
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekir
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
 DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
 A:Reference number: A71000; MUID:98344137
 A:Accession: A71067
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-483 <RAW>
 A:Cross-references: GB:AP000005; NID:g3236132; PID:d1031274; PID:g3257648
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH1231

Query Match 52.6%; Score 41; DB 2; Length 483;
 Best Local Similarity 58.3%; Pred. No. 10;
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 AITKLKLEGFDT 15
 ||| ||| |||
 Db 348 AVFKMELEGFET 359

RESULT 9

JX0629 formate dehydrogenase (EC 1.2.1.2) N (nitrate-inducible) beta chain - Escherichia coli
 N:Alternate names: formate dehydrogenase N iron-sulfur protein
 C:Species: Escherichia coli
 C>Date: 31-Mar-1992 #sequence_revision 17-Sep-1997 #text_change 05-Dec-1998
 C:Accession: F64900; JX0629
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 A:Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: F64900
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-294 <BLAT>

A:Cross-references: GB:AE000244; GB:U00096; NID:g1787742; PID:g1787749; UWGP:b1475
 A:Experimental source: strain K-12, substrain MG1655
 R:Berg, B.L.; Li, J.; Heider, J.; Stewart, V.
 J. Biol. Chem. 266, 22380-22385, 1991

A:Title: Nitrate-inducible formate dehydrogenase in Escherichia coli K-12. I. Nucleot
 A:Reference number: JX0628; MUID:92042178
 A:Accession: JX0629
 A:Molecule type: DNA

A:Residues: 1-100, 'ID', 103-294 <BER>

A:Cross-references: GB:M75029

A:Experimental source: strain K12

C:Genetics:

A:Gene: fdnH

C:Complex: heterotrimer; alpha, beta and gamma chains

C:Function:

A:Description: catalyzes the reversible oxidation of formate to carbon dioxide with N

A:Pathway: anaerobic nitrate respiration

A:Note: this form of the enzyme is induced by anaerobiosis and nitrate

A:Note: beta chain serves as electron transfer unit

C:Superfamily: ferredoxin 2[4Fe-4S] homology

C:Keywords: 4Fe-4S; heterotrimer; iron-sulfur protein; metalloprotein; NAD: oxidoredu

F:32-121/Domain: ferredoxin 2[4Fe-4S] homology #status atypical <FER1>

F:126-187/Domain: ferredoxin 2[4Fe-4S] homology #status atypical <FER2>

F:39,42,45,112/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:49,100,103,108/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:133,136,139,179/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:143,160,163,175/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 51.3%; Score 40; DB 2; Length 294;
 Best Local Similarity 50.0%; Pred. No. 9;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEQAITKLKLEGFDT 14
 ||| ||| |||
 Db 196 AEQRAVKLRGVE 209

RESULT 10

E71330 probable preprotein translocase subunit (seca) - syphilis spirochete
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
 C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
 C:Accession: E71330

R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khaliak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of *Treponema pallidum*, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770

A:Accession: E71330
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-916 <COL>
A:Cross-references: GB:AE001217; GB:AE000520; NID:g3322656; PID:g3322659
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0379
C:Superfamily: preprotein translocase secA; DEAD/H box helicase homology

Query Match 50.0%; Score 39; DB 2; Length 916;
Best Local Similarity 70.0%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 ITKLKLEGFD 14

||: |||||

Db 821 LTEYKLEGFD 830

RESULT 11
C64821
conserved hypothetical protein b0835 - *Escherichia coli*
C:Species: *Escherichia coli*
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 24-Mar-1999
C:Accession: C64821
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617

A:Accession: C64821
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-441 <BLAT>

A:Cross-references: GB:AE000185; GB:U00096; NID:g1787047; PID:g1787057; UWGP:b0835
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: conserved hypothetical protein b0835

Query Match 50.0%; Score 39; DB 2; Length 441;
Best Local Similarity 42.9%; Pred. No. 21;
Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEQAITKLEGGFD 14

||: |||||

Db 24 SERILTELRTEGYD 37

RESULT 12
S51623
cut14 protein - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 04-Sep-1998
C:Accession: S51623
R;Saka, Y.; Saitoh, T.; Yamashita, Y.; Saitoh, S.; Takeuchi, M.; Nakaseko, Y.; Yanagida,
EMBO J. 13, 4938-4952, 1994
A:Title: Fission yeast cut3 and cut14, members of a ubiquitous protein family, are requi
A:Reference number: S51622; MUID:95045386
A:Accession: S51623

A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-1172 <SAK>

A:Cross-references: EMBL:D30787; NID:g577658; PID:d1007024; PID:g603500
C:Superfamily: conserved hypothetical P115 protein

Query Match 50.0%; Score 39; DB 2; Length 1172;
Best Local Similarity 50.0%; Pred. No. 60;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 AEQAITKLEGGFD 16

||: |||||

Db 414 AEKETNRLKLEGLNKQ 429

RESULT 13
C70814
hypothetical protein RV0851c - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 12-Feb-1999
C:Accession: C70814
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Felitwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: C70814

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-275 <COL>

A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PID:e1253993; PID:g291591
A:Experimental source: strain H37RV
C:Genetics:

A:Gene: RV0851c

C:Superfamily: short-chain alcohol dehydrogenase homology
F:7-187/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 50.0%; Score 39; DB 2; Length 275;
Best Local Similarity 58.3%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 QAITKLEGGFD 14

||: |||||

Db 45 QAVNHLRAEGFD 56

RESULT 14
C34106
protein kinase (EC 2.7.1.37), cGMP-dependent 2, type 2 - fruit fly (*Drosophila melano*
C:Species: *Drosophila melanogaster*
C:Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 23-Feb-1997
C:Accession: C34106
R;Kalderson, D.; Rubin, G.M.
J. Biol. Chem. 264, 10738-10748, 1989

A:Title: cGMP-dependent protein kinase genes in *Drosophila*.
A:Reference number: A34106; MUID:89278147
A:Accession: C34106

A:Status: preliminary

A:Molecule type: DNA; mRNA

A:Residues: 1-575 <KAL>

A:Cross-references: GB:M30147; GB:J04817

C:Genetics:

A:Gene: FlyBase:for

A:Cross-references: FlyBase:FBgn0000721

C:Superfamily: cAMP receptor protein cyclic nucleotide-binding domain homology
C:Keywords: cGMP binding; phosphotransferase; serine/threonine-specific protein kinas
F:326-443/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CA
F:444-568/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CA

Query Match 50.0%; Score 39; DB 2; Length 575;
Best Local Similarity 81.8%; Pred. No. 28;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EQAITKLEGG 12

|||||

```

Db      291 EQAIEGLKLEG 301

RESULT  15
D34106
protein kinase (EC 2.7.1.37), cGMP-dependent 2, type cD5 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 10-Oct-1997
C:Accession: D34106
R:Kalderon, D.; Rubin, G.M.
J. Biol. Chem. 264, 10738-10748, 1989
A:Title: cGMP-dependent protein kinase genes in Drosophila.
A:Reference number: A34106; MUID:89278147
A:Accession: D34106
A:Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-614 <KAL>
A:Cross-references: GB:M27122; GB:M27123
C:Genetics:
A:Gene: FlyBase:for
A:Cross-references: FlyBase:FBgn0000721
C:Superfamily: cAMP receptor protein cyclic nucleotide-binding domain homology
C:Keywords: cGMP binding; phosphotransferase; serine/threonine-specific protein kinase
F:326-482/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP1>
F:483-607/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <CAP2>

Query Match      50.0%; Score 39; DB 2; Length 614;
Best Local Similarity 81.8%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 EQAITKLEKLEG 12
      |||| |||||
Db      291 EQAIEGLKLEG 301

```

Search completed: November 13, 1999, 12:07:57
 Job time: 2069 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:33:31 ; Search time 51.07 Seconds
(without alignments)
8.856 Million cell updates/sec

Title: US-08-913-430-5
Perfect score: 78
Sequence: 1 AQAATKCLKGFDTQ 16

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	78	100.0	416	1	P46_MYCHY
2	41	52.6	1465	1	AT7B_HUMAN
3	41	52.6	1451	1	AT7B_RAT
4	41	52.6	92	1	IC11_PHAAN
5	41	52.6	501	1	SYK_CAMJE
6	40	51.3	294	1	FDNH_ECOLI
7	39	50.0	1172	1	CUI4_SCHPO
8	39	50.0	933	1	KGP3_DREMA
9	39	50.0	916	1	SECA_TREPA
10	39	50.0	441	1	YLIG_ECOLI
11	38	48.7	3567	1	ERV2_SACER
12	38	48.7	226	1	RNC_ECOLI
13	38	48.7	541	1	SR34_YEAST
14	38	48.7	259	1	TERA_BPRD
15	38	48.7	145	1	YBCH_ECOLI
16	38	48.7	607	1	YK28_ARCFU
17	38	48.7	257	1	YXJF_BACSU
18	37	5	319	1	FTSN_ECOLI
19	37	47.4	227	1	CPAR_HAEIN
20	37	47.4	562	1	HMD2_ARATH
21	37	47.4	482	1	RS1_MYCLE
22	37	47.4	485	1	SR54_SYN7
23	37	47.4	847	1	STAG_HUMAN
24	37	47.4	109	1	YCA6_YEAST
25	37	47.4	461	1	YMBF_YEAST
26	37	47.4	150	1	YXBA_BACSU
27	36	5	3947	1	SID2_USTMA
28	36	46.8	102	1	ACYM_ANAPL
29	36	46.2	102	1	ACYM_CHICK
30	36	46.2	102	1	ACYM_MELGA
31	36	46.2	286	1	CAS1_CHICK
32	36	46.2	284	1	CAS1_MOUSE
33	36	46.2	734	1	MCW5_HUMAN
34	36	46.2	733	1	MCW5_MOUSE
35	36	46.2	426	1	MOEA_MYCTU
36	36	46.2	2411	1	MYSA_DROME
37	36	46.2	808	1	SECA_MYCPN
38	36	46.2	702	1	SPOT_ECOLI
39	36	46.2	427	1	Y428_METJA
40	36	46.2	193	1	YEXN_AERSA
41	36	46.2	1037	1	Y0J8_YEAST
42	35	44.9	357	1	AMRP_HUMAN
43	35	44.9	503	1	C7B6_ARATH

44 35 44.9 364 1 FTZ1_METJA Q57816 methanococ
45 35 44.9 417 1 YOR2_NEIGO O33369 neisseria g

ALIGNMENTS

```
RESULT 1
P46_MYCHY 1
AC P46_MYCHY STANDARD; PRT; 416 AA.
AC P46192;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 46 KD SURFACE ANTIGEN PRECURSOR (P46).
OS MYCOPLASMA HYOPNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 67-74 AND 338-358.
RC STRAIN-ATCC 25934 / J;
RX MEDLINE; 95204368.
RA FUTO S., SETO Y., MITSUSE S., MORI Y., SUZUKI T., KAWAI K.;
RT "Molecular cloning of a 46-kilodalton surface antigen (P46) gene from
RT Mycoplasma hyopneumoniae: direct evidence of CGG codon usage for
RT arginine."
RL J. BACTERIOL. 177:1915-1917(1995).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID ANCHOR
CC (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16682; G1220114; ALT INIT.
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
CC PFAM: PF00532; Peripla_BP_like; 1.
CC ANTIGEN: MEMBRANE; LIPOPROTEIN; SIGNAL.
CC SIGNAL 1 27 POTENTIAL.
CC FT CHAIN 28 416 46 KD SURFACE ANTIGEN.
CC FT LIPID 28 28 N-ACYL DIGLYCERIDE (POTENTIAL).
CC SQ SEQUENCE 416 AA; 45401 MW; 3EA04ED CRC32;

Query Match 100.0%; Score 78; DB 1; Length 416;
Best Local Similarity 100.0%; Pred. No. 2e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQAATKCLKGFDTQ 16
Db 298 AQAATKCLKGFDTQ 313
|||||
Db 298 AQAATKCLKGFDTQ 313

RESULT 2
AT7B_HUMAN
ID AT7B_HUMAN STANDARD; PRT; 1465 AA.
AC P35670; Q16318; Q16319;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE COPPER-TRANSPORTING ATPASE 2 (EC 3.6.1.36) (COPPER PUMP 2) (WILSON
DE DISEASE-ASSOCIATED PROTEIN).
GN ATP7B OR WND OR PWD OR WCL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94129610.
```

RA PETRUKHIN K., FISCHER S.G., PIRASTU M., TANZI R.E., CHERNOV I.,
RA DEVOTO M., BRZUSTOWICZ L.M., CANYANIS E., VITALE E., RUSSO J.J.,
RA MATSEANE D., BOUKKALTER B., WASCO W., FIGUS A.L., LOUDIANOS G.,
RA CAO A., STERNLIEB I., EVGRAFOV O., PARANO E., PAVONE L., WARBURTON D.,
RA OTT J., PENCHASZADEH G.K., SCHEINBERG I.H., GILLIAM T.C.;
RT "Mapping, cloning and genetic characterization of the region
RT containing the Wilson disease gene.";
RL NAT. GENET. 5:338-343(1993).
RN [2]
RP SEQUENCE OF 33-1465 FROM N.A.
RX MEDLINE; 94129609.
RA BULL P.C., THOMAS G.R., ROMMENS J.M., FORBES J.R., COX D.W.;
RT "The Wilson disease gene is a putative copper transporting P-type
RT ATPase similar to the Menkes gene.";
RL NAT. GENET. 5:327-337(1993).
RN [3]
RP PARTIAL SEQUENCE FROM N.A., AND VARIANTS WD GLN-1069 AND SER-1270.
RX MEDLINE; 94129611.
RA TANZI R.E., PETRUKHIN K., CHERNOV I., PELLEQUER J.L., WASCO W.,
RA ROSS B., ROMANO D.M., PARANO E., PAVONE L., BRZUSTOWICZ L.M.,
RA DEVOTO M., PEPPERCORN J., BUSH A.I., STERNLIEB I., PIRASTU M.,
RA GUSELLA J.F., EVGRAFOV O., PENCHASZADEH G.K., HONIG B., EDELMAN I.S.,
RA SOARES M.B., SCHEINBERG I.H., GILLIAM T.C.;
RT "The Wilson disease gene is a copper transporting ATPase with
RT homology to the Menkes disease gene.";
RL NAT. GENET. 5:344-350(1993).
RN [4]
RP SEQUENCE OF 489-830 FROM N.A.
RX MEDLINE; 94071954.
RA YAMAGUCHI Y., HEINY M.E., GITLIN J.D.;
RT "Isolation and characterization of a human liver cDNA as a candidate
RT gene for Wilson disease.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 197:271-277(1993).
RN [5]
RP VARIANTS WD, AND POLYMORPHISM.
RX MEDLINE; 96065019.
RA FIGUS A., ANGIUS A., LOUDIANOS G., BERTINI C., DESSI V., LOI A.,
RA DEIANA M., LOVICU M., OLLA N., SOLE G., DE VIRGILIIS S., LILLIU F.,
RA FARCI A.M.G., NURCHI A., GIACCHINO R., BARABINO A., MARAZZI M.,
RA ZANCAN L., GREGGIO N.A., MACCELLINI M., SOLINAS A., DEPIANO A.,
RA BARBERA C., DEVOTO M., OSOILU S., KOCAR N., AKAR N., KARAYALCIN S.,
RA MOKINI V., CULLUFI P., BALESTRIERI A., CAO A., PIRASTU M.;
RT "Molecular pathology and haplotype analysis of Wilson disease in
RT Mediterranean populations.";
RL AM. J. HUM. GENET. 57:1318-1324(1995).
RN [6]
RP VARIANTS WD, AND POLYMORPHISM.
RX MEDLINE; 95235569.
RA THOMAS G.R., FORBES J.R., ROBERTS E.A., WALSH J.M., COX D.W.;
RT "The Wilson disease gene: spectrum of mutations and their
RT consequences.";
RL NAT. GENET. 9:210-216(1995).
RN [7]
RP VARIANTS WD.
RX MEDLINE; 97085557.
RA LOUDIANOS G., DESSI V., ANGIUS A., LOVICU M., LOI A., DEIANA M.,
RA AKAR N., VAJRO P., FIGUS A., CAO A., PIRASTU M.;
RT "Wilson disease mutations associated with uncommon haplotypes in
RT Mediterranean patients.";
RL HUM. GENET. 98:640-642(1996).
RN [8]
RP VARIANTS WD GLN-778 AND LEU-778.
RX MEDLINE; 96375768.
RA CHUANG L.-M., WU H.-P., JANG M.-H., WANG T.-R., SUE W.-C., LIN B.-J.,
RA COX D.W., TAI T.-Y.;
RT "High frequency of two mutations in codon 778 in exon 8 of the ATP7B
RT gene in Taiwanese families with Wilson disease.";
RL J. MED. GENET. 33:521-523(1996).
RN [9]
RP VARIANTS WD.
RX MEDLINE; 97456422.
RA SHAH A.B., CHERNOV I., ZHANG H.T., ROSS B.M., DAS K., LUTSENKO S.,

RA PARANO E., PAVONE L., EVGRAFOV O., IVANOVA-SMOLENSKAYA I.A.,
RA ANNEREN G., WESTERMARK K., URRUTIA F.H., PENCHASZADEH G.K.,
RA STERNLIEB I., SCHEINBERG I.H., GILLIAM T.C., PETRUKHIN K.;
RT "Identification and analysis of mutations in the Wilson disease gene
RT (ATP7B): population frequencies, genotype-phenotype correlation, and
RT functional analyses.";
RL AM. J. HUM. GENET. 61:317-328(1997).
RN [10]
RP VARIANT WD GLY-1285--ILE-1292 DEL, AND VARIANT VAL-1278.
RX MEDLINE; 97365940.
RA ORRU S., THOMAS G., LOIZEDDA A., COX D.W., CONTU L.;
RT "24 bp deletion and Ala1278 to Val mutation of the ATP7B gene in a
RT Sardinian family with Wilson disease.";
RL HUM. MUTAT. 10:84-85(1997).
RN [11]
RP VARIANTS WD, AND VARIANTS.
RX MEDLINE; 98141682.
RA KALINSKY H., FUNES A., ZELDIN A., PEL-OR Y., KOROSTISHEVSKY M.,
RA GERSHONI-BARUCH R., FARRER L.A., BONNE-TAMIR B.;
RT "Novel ATP7B mutations causing Wilson disease in several Israeli
RT ethnic groups.";
RL HUM. MUTAT. 11:145-151(1998).
RN [12]
RP VARIANTS WD L-778;V-874;F-1083, & VARIANTS R-832;I-864;M-1109;A-1140.
RX MEDLINE; 98213465.
RA KIM E.K., YOO O.J., SONG K.Y., YOO H.W., CHOI S.Y., CHO S.W.,
RA HAHN S.H.;
RT "Identification of three novel mutations and a high frequency of the
RT Arg778Leu mutation in Korean patients with Wilson disease.";
RL HUM. MUTAT. 11:275-278(1998).
RN [13]
RP VARIANTS WD LEU-778; VAL-874; GLY-919; SER-1186; ALA-1267 & SER-1270.
RX MEDLINE; 98112500.
RA YAMAGUCHI A., MATSUURA A., ARASHIMA S., KIKUCHI Y., KIKUCHI K.;
RT "Mutations of ATP7B gene in Wilson disease in Japan: identification
RT of nine mutations and lack of clear founder effect in a Japanese
RT population.";
RL HUM. MUTAT. SUPPL. 1:S320-S322(1998).
RN [14]
RP VARIANTS WD.
RX MEDLINE; 98334344.
RA LOUDIANOS G., DESSI V., LOVICU M., ANGIUS A., NURCHI A.,
RA STURNIOLO G.C., MARCELLINI M., ZANCAN L., BRAGETTI P., AKAR N.,
RA YAGGI R., VEGNENTE A., CAO A., PIRASTU M.;
RT "Further delineation of the molecular pathology of Wilson disease in
RT the Mediterranean population.";
RL HUM. MUTAT. 12:89-94(1998).
RN [15]
RP VARIANTS WD.
RX MEDLINE; 99045370.
RA TSAI C.-H., TSAI F.-J., WU J.-Y., CHANG J.-G., LEE C.-C., LIN S.-P.,
RA YANG C.-F., JONG Y.-J., LO M.-C.;
RT "Mutation analysis of Wilson disease in Taiwan and description of six
RT new mutations.";
RL HUM. MUTAT. 12:370-376(1998).
RN CC - FUNCTION: MAY FUNCTION IN THE EXPORT OF COPPER IN CELLS.
CC - CATALYTIC ACTIVITY: ATP + H(2)O - ADP + ORTHOPHOSPHATE.
CC - SUBUNIT: MONOMER.
CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC - TISSUE SPECIFICITY: MOST ABUNDANT IN THE LIVER AND KIDNEY.
CC - DISEASE: DEFECTS IN ATP7B ARE ASSOCIATED WITH WILSON DISEASE (WD);
CC AN AUTOSOMAL RECESSIVE DISORDER OF COPPER METABOLISM IN WHICH
CC COPPER CANNOT BE INCORPORATED INTO CERULOPLASMIN IN LIVER, AND
CC CANNOT BE EXCRETED FROM THE LIVER INTO THE BILE. COPPER
CC ACCUMULATES IN THE LIVER AND SUBSEQUENTLY IN THE BRAIN AND KIDNEY.
CC THE DISEASE IS CHARACTERIZED BY NEUROLOGIC MANIFESTATIONS AND
CC SIGNS OF CIRRHOSIS.
CC - SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES). SUBFAMILY 1B.
CC - SIMILARITY: CONTAINS 6 COPIES OF THE HEAVY-METAL-ASSOCIATED
CC (HMA) DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; U01700; G551502; -;
 DR EMBL; U03464; G434683; -;
 DR EMBL; L25591; G452075; ALT_SEQ.
 DR EMBL; L25591; G452076; ALT_SEQ.
 DR EMBL; L25442; G409358; ALT_SEQ.
 DR EMBL; S77446; E198378; ALT_SEQ.
 DR EMBL; S77447; G957353; ALT_SEQ.
 DR EMBL; S77450; G957355; ALT_SEQ.
 DR MIM; 277900; -;
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 DR PROSITE; PS01047; HMA; 6.
 DR PFAM; PF00122; E1-E2_ATPase; 1.
 DR PFAM; PF00403; HMA; 6.
 DR HSSP; Q04656; 2AWO.

KW HYDROLASE; COPPER TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
 KW ATP-BINDING; METAL-BINDING; COPPER; REPEAT; DISEASE MUTATION;
 KW POLYMORPHISM.

FT DOMAIN 1 649 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 650 673 POTENTIAL.
 FT TRANSMEM 685 713 POTENTIAL.
 FT TRANSMEM 728 754 POTENTIAL.
 FT TRANSMEM 763 781 POTENTIAL.
 FT TRANSMEM 921 949 POTENTIAL.
 FT TRANSMEM 964 991 POTENTIAL.
 FT TRANSMEM 1036 1069 POTENTIAL.
 FT TRANSMEM 1177 1197 POTENTIAL.
 FT TRANSMEM 1322 1372 POTENTIAL.
 FT DOMAIN 64 93 HMA 1.
 FT DOMAIN 149 178 HMA 2.
 FT DOMAIN 263 292 HMA 3.
 FT DOMAIN 365 394 HMA 4.
 FT DOMAIN 494 523 HMA 5.
 FT DOMAIN 570 599 HMA 6.

Query Match 52.6%; Score 41; DB 1; Length 1465;
 Best Local Similarity 43.8%; Pred. No. 24;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEOAITKLEGGFTQ 16
 |:::|:::|
 Db 140 AOEAVKLRVEGTCQ 155

RESULT 3
 ID AT7B_RAT STANDARD; PRT; 1451 AA.
 AC Q64535;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE COPPER-TRANSPORTING ATPASE 2 (EC 3.6.1.36) (COPPER PUMP 2).
 GN ATP7B.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=LIVER;
 RX MEDLINE; 95038843.
 RA WU J., FORBES J.R., CHEN H.S., COX D.W.;
 RT "The LEC rat has a deletion in the copper transporting ATPase gene homologous to the Wilson disease gene."
 RL NAT. GENET. 7:541-545(1994).
 CC -1- FUNCTION: MAY FUNCTION IN THE EXPORT OF COPPER IN CELLS.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O -> ADP + ORTHOPHOSPHATE.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2 ATPASES), SUBFAMILY IB.
 CC -1- SIMILARITY: CONTAINS 6 COPIES OF THE HEAVY-METAL-ASSOCIATED (HMA) DOMAIN.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U08344; G555676; -;
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 DR PROSITE; PS01047; HMA; 5.
 DR PFAM; PF00122; E1-E2_ATPase; 1.
 DR PFAM; PF00403; HMA; 5.
 DR HSSP; Q04656; 2AWO.

KW HYDROLASE; COPPER TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
 KW ATP-BINDING; METAL-BINDING; COPPER; REPEAT.
 KW CYTOPLASMIC (POTENTIAL).

FT DOMAIN 1 642 POTENTIAL.
 FT TRANSMEM 643 666 POTENTIAL.
 FT TRANSMEM 678 706 POTENTIAL.
 FT TRANSMEM 721 747 POTENTIAL.
 FT TRANSMEM 756 774 POTENTIAL.
 FT TRANSMEM 914 942 POTENTIAL.
 FT TRANSMEM 957 984 POTENTIAL.
 FT TRANSMEM 1029 1062 POTENTIAL.
 FT TRANSMEM 1165 1185 POTENTIAL.
 FT TRANSMEM 1310 1358 POTENTIAL.
 FT DOMAIN 63 92 HMA 1.
 FT DOMAIN 148 177 HMA 2.
 FT DOMAIN 262 291 HMA 3.
 FT DOMAIN 361 390 HMA 4.
 FT DOMAIN 487 516 HMA 5.
 FT DOMAIN 563 592 HMA 6.
 FT MOD_RES 1020 1020 PHOSPHORYLATION (PROBABLE).
 FT CARBOHYD 885 885 POTENTIAL.
 SQ SEQUENCE 1451 AA; 155990 MW; B90A8FC3 CRC32;

Query Match 52.6%; Score 41; DB 1; Length 1451;
 Best Local Similarity 43.8%; Pred. No. 24;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEOAITKLEGGFTQ 16
 |:::|:::|
 Db 139 AOEAVKLRVEGTCQ 154

RESULT 4
 ID IC11_PHAAN STANDARD; PRT; 92 AA.
 AC P16064;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
 DE SUBTILISIN INHIBITORS I AND II (ASI-I AND ASI-II).
 OS PHASEOLUS ANGULARIS (ADZUKI BEAN) (VIGNA ANGULARIS).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC FABALES; FABACEAE; PAPILIONOIDEAE; VIGNA.
 RN [1]
 RP SEQUENCE.
 RX MEDLINE; 90186673.
 RA NOZAWA H., YAMAGATA H., AIZONO Y., YOSHIKAWA M., IWASAKI T.;
 RT "The complete amino acid sequence of a subtilisin inhibitor from adzuki beans (Vigna angularis).";
 RL J. BIOCHEM. 106:1003-1008(1989).
 CC -1- FUNCTION: INHIBITOR OF SUBTILISIN.
 CC -1- SIMILARITY: BELONGS TO THE POTATO CHYMOTRYPSIN INHIBITOR I FAMILY


```
CC OF SERINE PROTEASE INHIBITOR.
DR PIR: JX0089; JX0089.
DR PROSITE: PS00285; POTATO_INHIBITOR; 1.
DR PRAM: PF00280; potato_inhib; 1.
DR HSP: P01053; ICIR.
KW SERINE PROTEASE INHIBITOR.
FT CHAIN 1 92 ASI-I.
FT CHAIN 20 92 ASI-II.
FT ACT SITE 68 69 REACTIVE BOND.
SQ SEQUENCE 92 AA; 10384 MW; FCF7175A CRC32;

Query Match 52.68; Score 41; DB 1; Length 92;
Best Local Similarity 62.58; Pred. No. 1.4;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AEQAITKLEGFDTQ 16
DB 41 AEQAETKIKEMVDVQ 56

RESULT 5
SYK_CAMJE STANDARD; PRT; 501 AA.
AC P41258;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE LYSYL-TRNA SYNTHETASE (EC 6.1.1.6) (LYSINE-TRNA LIGASE) (LYSRS).
GN LYSS.
OS CAMPYLOBACTER JEJUNI.
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;
ON CAMPYLOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 43431 / TGH9011;
RX MEDLINE; 92121106.
RA CHAN V.L., BINGHAM H.;
RT "Lysyl-trna synthetase gene of Campylobacter jejuni.";
RL J. BACTERIOL. 174:695-701(1992).
CC -1- CATALYTIC ACTIVITY: ATP + L-LYSINE + TRNA(LYS) -> AMP +
CC PHOSPHATE + L-LYSYL-TRNA(LYS).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63448; G144208;
DR PIR; A42609; A42609.
DR PROSITE; PS00179; AA-TRNA_LIGASE_II_1; 1.
DR PRAM; PS00339; AA-TRNA_LIGASE_II_2; 1.
DR HSP; P14825; LYL1.
DR AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
SQ SEQUENCE 501 AA; 57859 MW; 1BF07DAF CRC32;

Query Match 52.68; Score 41; DB 1; Length 501;
Best Local Similarity 53.88; Pred. No. 8;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 EQAITKLEGFDTQ 14
DB 338 EKILTKLADGFE 350

us-08-913-430-5.rsp
```

```
RESULT 6
FDNH_ECOLI STANDARD; PRT; 294 AA.
AC P24184; P77166;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE FORMATE DEHYDROGENASE, NITRATE-INDUCIBLE, IRON-SULFUR SUBUNIT (FORMATE
DE DEHYDROGENASE-N BETA SUBUNIT) (FDH-N BETA SUBUNIT) (ANAEROBIC FORMATE
DE DEHYDROGENASE IRON-SULFUR SUBUNIT).
GN FDNH.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
ON ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 92042178.
RA BERG B.L., LI J., HEIDER J., STEWART V.;
RT "Nitrate-inducible formate dehydrogenase in Escherichia coli K-12. I.
RT Nucleotide sequence of the fdnH1 operon and evidence that opal (UGA)
RT encodes selenocysteine."
RL J. BIOL. CHEM. 266:22380-22385(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASSNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL SCIENCE 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANAI K., KASAI H.,
RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,
RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,
RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: FORMATE DEHYDROGENASE ALLOWS E. COLI TO USE FORMATE AS
CC MAJOR ELECTRON DONOR DURING ANAEROBIC RESPIRATION, WHEN NITRATE IS
CC USED AS ELECTRON ACCEPTOR. THE BETA CHAIN IS AN ELECTRON TRANSFER
CC UNIT CONTAINING 4 CYSTEINE CLUSTERS INVOLVED IN THE FORMATION OF
CC IRON-SULFUR CENTRES. ELECTRONS ARE TRANSFERRED FROM THE GAMMA
CC CHAIN TO THE MOLYBDENUM COFACTOR OF THE ALPHA SUBUNIT (BY
CC SIMILARITY).
CC -1- PATHWAY: ANAEROBIC NITRATE RESPIRATION.
CC -1- SUBUNIT: FORMATE DEHYDROGENASE IS A MEMBRANE-BOUND COMPLEX, FORMED
CC BY SUBUNITS ALPHA, BETA AND GAMMA.
CC -1- INDUCTION: BY NITRATE, UNDER ANAEROBIC CONDITIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M75029; -; NOT_ANNOTATED_CDS.
DR EMBL; AE000244; G1787749; -
DR EMBL; D90788; G1742408; -
DR EMBL; D90789; G1742418; -
DR PIR; JS0629; JS0629.
DR ECGENE; EG11228; FDNH.
DR PROSITE; PS00198; 4FE4S-FERREDOXIN; 1.
DR PFAM; PF00037; fer4; 2.
DR ELECTRON TRANSPORT; 4FE-4S; IRON-SULFUR; TRANSMEMBRANE.
KW
```

FT METAL 39 39 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 42 42 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 45 45 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 49 49 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
 FT METAL 100 100 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 103 103 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 108 108 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 112 112 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
 FT METAL 133 133 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 136 136 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 139 139 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 143 143 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
 FT METAL 160 160 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 163 163 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 175 175 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT METAL 179 179 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
 FT TRANSMEM 259 281 POTENTIAL.
 FT CONFLICT 101 102 MH -> ID (IN REF. 1).
 SQ SEQUENCE 294 AA; 32238 MW; 0C4326B6 CRC32;

Query Match 51.3%; Score 40; DB 1; Length 294;
 Best Local Similarity 50.0%; Pred. No. 7;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEQAITKLEGGD 14
 ||| : ||| : ||| :
 Db 196 AEQVAKLKARGYE 209

RESULT 7
 CUL14_SCHPO STANDARD; PRT; 1172 AA.
 AC P41003;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE CHROMOSOME SEGREGATION PROTEIN CUT14.
 GN CUT14.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIAZCOMYCETES;
 OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
 OC SCHIZOSACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95045386.
 RA SAKA Y., SUTANI T., YAMASHITA Y., SAITOH S., TAKEUCHI M.,
 RA NAKASEKO Y., YANAGIDA M.;
 RT "Fission yeast cut3 and cut14, members of a ubiquitous protein
 family, are required for chromosome condensation and segregation in
 mitosis."
 RL EMBO J. 13:4938-4952(1994).
 CC -1- FUNCTION: REQUIRED FOR CHROMOSOME CONDENSATION AND SEGREGATION
 CC IN MITOSIS
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D30787; G503500;
 KW MITOSIS; ATP-BINDING; COILED COIL; NUCLEAR PROTEIN.
 FT NP_BIND 32 39 ATP (POTENTIAL).
 FT DOMAIN 172 204 COILED COIL (POTENTIAL).
 FT DOMAIN 258 507 COILED COIL (POTENTIAL).
 FT DOMAIN 676 941 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1172 AA; 134327 MW; D261ABE3 CRC32;

Query Match 50.0%; Score 39; DB 1; Length 1172;
 Best Local Similarity 50.0%; Pred. No. 43;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 AEQAITKLEGGDFTQ 16
 ||| : ||| : ||| :
 Db 414 AEKTRRLKLEGLNQ 429

RESULT 8
 KGP3_DRONE STANDARD; PRT; 933 AA.
 ID KGP3_DRONE
 AC P32023;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE CGMP-DEPENDENT PROTEIN KINASE, ISOZYME 2 FORMS T2/CD5 (EC 2.7.1.-)
 DE (CGK) (PROTEIN FORAGING).
 GN FOR OR PKG24A OR PKG2 OR DG2.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 89325663.
 RA KALDERON D., RUBIN G.M.;
 RT "The cDNA of the two isoforms of bovine CGMP-dependent protein
 kinase."
 RL J. BIOL. CHEM. 264:10738-10748(1989).
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN -> ADP + A PHOSPHOPROTEIN.
 CC -1- THIS KINASE CONTAINS THE FOLLOWING DOMAINS: TWO CGMP-BINDING
 CC REGIONS (THE FIRST BEING TRUNCATED) AND THE CATALYTIC DOMAIN.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST FOUR FORMS OF THIS KINASE ARE
 CC PRODUCED BY ALTERNATIVE SPLICING. THE SEQUENCE IS SHOWN HERE
 CC IS CALLED T2.
 CC -1- SIMILARITY: TO OTHER CGMP-DEPENDENT PROTEIN KINASES, AND ALSO
 CC TO CAMP-DEPENDENT PROTEIN KINASES.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; M30147; G157214;
 DR EMBL; M30148; G157222;
 DR PIR; C34106; C34106.
 DR PIR; D34106; D34106.
 DR FLYBASE; FBgn0000721; for.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00888; CNMP_BINDING_1; 1.
 DR PROSITE; PS00889; CNMP_BINDING_2; 2.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS50042; CNMP_BINDING_3; 2.
 DR PFAM; PF00027; CNMP_binding; 2.
 DR PFAM; PF00069; pkinase; 1.
 DR PFAM; PF00433; pkinase_C; 1.
 DR HSP; P05132; 2CPK.
 KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
 CCMP-BINDING; ALTERNATIVE SPLICING.
 FT NP_BIND 408 490 CGMP 1 (TRUNCATED) (BY SIMILARITY).
 FT NP_BIND 491 606 CGMP 2 (BY SIMILARITY).
 FT DOMAIN 622 881 PROTEIN KINASE.
 FT NP_BIND 628 636 ATP (BY SIMILARITY).
 FT BINDING 652 652 ATP (BY SIMILARITY).
 FT ACT_SITE 746 746 BY SIMILARITY.
 FT VARSPPLIC 368 368 G -> D (IN T2).
 FT VARSPPLIC 369 407 MISSING (IN T2).

10

CC -!- SIMILARITY: BELONGS TO THE RNASE III FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; X02946; G42769; -
 DR EMBL; X02673; G42766; -
 DR EMBL; D60474; G987642; -
 DR EMBL; U36841; G1033156; -
 DR EMBL; AE000343; G1788920; -
 DR EMBL; M26415; G499371; -
 DR EMBL; M14558; G145858; -
 DR PIR; A26023; NREC3V.
 DR PIR; A24022; NREC3.
 DR ECOGENE; EG10857; RNC.
 DR PROSITE; PS00517; RIBONUCLEASE_III; 1.
 DR PFAM; PF00035; dsrm; 1.
 DR PFAM; PF00636; Ribonuclease_3; 1.
 DR HYDROLASE; NUCLEASE; ENDONUCLEASE; RNA-BINDING.
 KW DOMAIN 207 223
 FT MUTAGEN 44 44 G->D: LOSS OF ACTIVITY (MUTANT RNC-105).
 FT CONFLICT 168 195 HLPPTLVVGVRCGEAHQDEFTICQVS -> PSAADLISG
 FT SPTGWSKRTINLLSTARSV (IN REF. 2).
 SQ SEQUENCE 226 AA; 25550 MW; D5B8B619 CRC32;

Query Match 48.7%; Score 38; DB 1; Length 226;
 Best Local Similarity 72.7%; Pred. No. 12;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AEQAITKKLE 11
 |||: |||:
 Db 216 AEQALKLELE 226

RESULT 13
 SR54_YEAST
 ID SR54_YEAST STANDARD; PRT; 541 AA.
 AC P20424;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE SIGNAL RECOGNITION PARTICLE 54 KD PROTEIN HOMOLOG (SRP54).
 GN SRP54 OR SRH1 OR YPR088C OR P9513.14.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETACEAE; SACCHAROMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AH22;
 RX MEDLINE; 90256724.
 RA MAYA Y., NAKANO A., ITO K., MORI M.;
 RT "Isolation of a yeast gene, SRH1, that encodes a homologue of the 54K
 RT subunit of mammalian signal recognition particle.";
 RL J. BIOCHEM. 107:457-463(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 90094527.
 RA HANN B.C., PORITZ M.A., WALTER P.;
 RT "Saccharomyces cerevisiae and Schizosaccharomyces pombe contain a
 RT homologue to the 54-kD subunit of the signal recognition particle
 RT that in S. cerevisiae is essential for growth.";
 RL J. CELL BIOL. 109:3223-3230(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
 RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J.,

RA KUCABA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M.,
 RA JOHNSON D., JOHNSON L., LANGSTON M., MILLER N., NHAN M., PAULEY A., PELUSO D.,
 RA MARDIS E., MENEZES S., MILLER N., NHAN M., PAULEY A., PELUSO D.,
 RA RIEKEN L., RILES L., TAICH A., TREVASKIS E., VIGNATI D.,
 RA WILSON L., WOLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- FUNCTION: BINDS TO THE SIGNAL SEQUENCE OF PRESECTORY PROTEIN
 CC WHEN THEY EMERGE FROM THE RIBOSOMES AND TRANSFERS THEM TO TRAM
 CC (TRANSLOCATING CHAIN-ASSOCIATING MEMBRANE PROTEIN).
 CC -!- SUBUNIT: YEAST SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA
 CC MOLECULE (SCR1) AND AT LEAST SEVEN PROTEIN SUBUNITS: SRP72, SRP68,
 CC SRP54, SEC55, SRP21, SPR14 AND SRP7.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- DOMAIN: HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN BINDS GTP: THE
 CC M-DOMAIN BINDS THE 7S RNA IN PRESENCE OF SRP19 AND ALSO BINDS THE
 CC SIGNAL SEQUENCE.
 CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X16908; G4538; -
 DR EMBL; M55517; G172705; -
 DR EMBL; X51614; G4542; -
 DR EMBL; U51033; G1230693; -
 DR PIR; JX0112; JX0112.
 DR SGD; L0002064; SRP54.
 DR PROSITE; PS00300; SRP54; 1.
 DR PFAM; PF00448; SRP54; 1.
 DR HSSP; P10121; IFTS.
 DR KW SIGNAL RECOGNITION PARTICLE; GTP-BINDING; RNA-BINDING.
 FT DOMAIN 1 303
 FT G-DOMAIN
 FT NP_BIND 304 541
 FT NP_BIND 116 123 GTP (BY SIMILARITY).
 FT NP_BIND 198 202 GTP (BY SIMILARITY).
 FT NP_BIND 256 259 GTP (BY SIMILARITY).
 FT CONFLICT 136 136 R -> E (IN REF. 1).
 SQ SEQUENCE 541 AA; 59624 MW; 2BC8E27D CRC32;

Query Match 48.7%; Score 38; DB 1; Length 541;
 Best Local Similarity 57.1%; Pred. No. 29;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Oy 1 AEQAITKKLEGED 14
 ||: ||: ||:
 Db 180 AEEGINKKKEKED 193

RESULT 14
 TERM_BPPRD
 ID TERM_BPPRD STANDARD; PRT; 259 AA.
 AC P09009;
 DT 01-NOV-1988 (REL. 09, CREATED)
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
 DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
 DE DNA TERMINAL PROTEIN (PROTEIN P8).
 GN VII.
 OS BACTERIOPHAGE PRD1.
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TECTIVIRIDAE; TECTIVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 88067710.
 RA HSIEH J.-C., JUNG G., LEAVITT M.C., ITO J.;
 RT "Primary structure of the DNA terminal protein of bacteriophage
 RT PRD1.";
 RL NUCLEIC ACIDS RES. 15:8999-9009(1987).
 RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE; 88112855.
RA SAVILAHTI H., BAMFORD D.H.;
RT "The complete nucleotide sequence of the left very early region of
RT Escherichia coli bacteriophage PRD1 coding for the terminal protein
and the DNA polymerase.";
RL GENE 57:121-130(1987).
CC -1- FUNCTION: COVALENTLY LINKED TO THE 5' END OF THE GENOME. USED
CC AS A PRIMER IN DNA REPLICATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M59077; G215766; -;
DR EMBL; X06321; G15801; -;
DR EMBL; M22161; G215751; -;
DR PIR; S01613; TPBPPR.
KW EARLY PROTEIN: DNA REPLICATION; DNA PRIMING;
KW COVALENT PROTEIN-DNA LINKAGE.
SQ SEQUENCE 259 AA; 29514 MW; A9C580FB CRC32;

Query Match 48.7%; Score 38; DB 1; Length 259;
Best Local Similarity 53.3%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 EQAITKLKLEGFDQ 16
Db ||:|||||
145 EQWTELKAAAGFQLQ 159

RESULT 15
YBCH_ECOLI
ID YBCH_ECOLI STANDARD; PRT; 145 AA.
AC P37325;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 16.8 KD PROTEIN IN ENVY-NFRA INTERGENIC REGION PRECURSOR.
GN YBCH.
OS ESCHERICHIA COLI.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RA MEDLINE; 94042874.
RA KLINO D.R., SINGER M.S., ROTHMAN-DENES L.B.;
RT "Two overlapping genes encoding membrane proteins required for
RT bacteriophage N4 adsorption.";
RL J. BACTERIOL. 175:7081-7085(1993).
RN [2]
RP IDENTIFICATION.
RA RUDD K.E.;
RL UNPUBLISHED OBSERVATIONS (AUG-1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L16945; -; NOT_ANNOTATED_CDS.
DR ECGENE; EGI2448; YBCH.
KW HYPOTHETICAL PROTEIN; SIGNAL.
FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 145 HYPOTHETICAL PROTEIN YBCH.
SQ SEQUENCE 145 AA; 16760 MW; D97BD396 CRC32;

Query Match 48.7%; Score 38; DB 1; Length 145;
Best Local Similarity 54.5%; Pred. No. 7.7;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 5 ITKLEGFDT 15
Db ||:|||||
44 MSQRLQGFDT 54

Search completed: November 13, 1999, 10:33:33
Job time: 5183 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:24 ; Search time 139.86 Seconds
(without alignments)
7.041 Million cell updates/sec

Title: US-08-913-430-5

Perfect score: 78
Sequence: 1 AQAATKCLKLEGFTQ 16

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPTREMBL_10.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	55.1	550	2 Q9ZCT7	Q9ZCT7 rickettsia
2	41	52.6	483	1 O58952	O58952 pyrococcus
3	41	52.6	355	2 O68607	O68607 campylobact
4	41	52.6	1462	11 Q64446	Q64446 mus musculus
5	40	51.3	489	5 O18258	O18258 caenorhabdi
6	40	51.3	2561	5 O46025	O46025 caenorhabdi
7	39	50.0	275	2 O53863	O53863 mycobacteri
8	39	50.0	505	2 O68274	O68274 alcaligenes
9	39	50.0	934	5 Q24302	Q24302 drosophila
10	39	50.0	894	5 Q24303	Q24303 drosophila
11	39	50.0	1330	6 O97961	O97961 vulpes vulp
12	38.5	49.4	119	2 Q48977	Q48977 mycoplasma
13	38	48.7	222	1 Q29571	Q29571 archaeoglob
14	38	48.7	296	2 P77641	P77641 escherichia
15	38	48.7	188	2 Q9ZKP8	Q9ZKP8 helicobacte
16	38	48.7	440	2 Q9ZE78	Q9ZE78 rickettsia
17	38	48.7	1402	2 Q9Z7Y9	Q9Z7Y9 chlamydia p
18	38	48.7	1974	5 Q21000	Q21000 caenorhabdi
19	37	47.4	259	2 Q9Z869	Q9Z869 chlamydia p
20	37	47.4	1191	3 Q42765	Q42765 candida alb
21	37	47.4	366	5 Q20005	Q20005 caenorhabdi
22	37	47.4	507	10 Q22233	Q22233 arabidopsis
23	37	47.4	583	10 O81437	O81437 arabidopsis
24	37	47.4	587	10 Q9ZQ09	Q9ZQ09 arabidopsis
25	37	47.4	3054	12 O88507	O88507 tobacco etc
26	36.5	46.8	248	1 O59175	O59175 pyrococcus
27	36	46.2	132	2 Q45284	Q45284 bacillus li
28	36	46.2	375	2 O51376	O51376 borrelia bu
29	36	46.2	443	2 O51666	O51666 borrelia bu

30	36	46.2	504	2 O07811	O07811 mycobacteri
31	36	46.2	1177	2 O67226	O67226 aquifex aeo
32	36	46.2	450	2 O07752	O07752 mycobacteri
33	36	46.2	1295	2 O68900	O68900 escherichia
34	36	46.2	349	2 P73453	P73453 synecocyst
35	36	46.2	424	2 Q9ZSG0	Q9ZSG0 mycobacteri
36	36	46.2	519	2 Q9Z3H2	Q9Z3H2 synecococc
37	36	46.2	632	3 Q04311	Q04311 saccharomyc
38	36	46.2	2101	4 Q14981	Q14981 homo sapien
39	36	46.2	2115	4 Q14980	Q14980 homo sapien
40	36	46.2	1701	4 O75643	O75643 homo sapien
41	36	46.2	407	4 Q15434	Q15434 homo sapien
42	36	46.2	1324	4 O94884	O94884 homo sapien
43	36	46.2	163	5 Q24412	Q24412 drosophila
44	36	46.2	1647	5 O45000	O45000 caenorhabdi
45	36	46.2	1201	5 O18392	O18392 drosophila

ALIGNMENTS

```

RESULT 1
Q9ZCT7 ID Q9ZCT7 PRELIMINARY; PRT; 550 AA.
AC Q9ZCT7;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE 60 KD CHAPERONIN (PROTEIN CPN60) (HEAT SHOCK PROTEIN
DE 60).
GS RP626.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsia; Rickettsia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RA ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA SICHERTZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,
RA ERIKSSON A.S., WINKLER H.H., KUHLAND C.G.;
RT "The Genome Sequence of Rickettsia prowazekii and the Origin of
RT Mitochondria.";
RL Nature 396:133-140(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RA ANDERSSON S.G.E.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR ENBL; AJ235272; CAA15067.1; -
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW Chaperone; ATP-binding.
SQ SEQUENCE 550 AA; 58987 MW; 05026E5A CRC32;

```

Query Match 55.1%; Score 43; DB 2; Length 550;
Best Local Similarity 56.2%; Pred. No. 17;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 AQAATKCLKLEGFTQ 16
| | : | | | : | |
Db 422 ASQTLTKLKVENDQQ 437

RESULT 2
O58952 ID O58952 PRELIMINARY; PRT; 483 AA.
AC O58952;

DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
 DE 483AA LONG HYPOTHETICAL PROTEIN.
 GN PH1231.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OT3;
 RX MEDLINE: 98344137.
 RA KAWABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,
 RA SARAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
 RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIYUZA H.,
 RA "KUCHI H.;
 RT "Complete sequence and gene organization of the genome of a
 RT hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA Res. 5:55-76(1998).
 DR EMBL: AP000005; BAA30331.1; -. 1372ED3D CRC32;
 SQ SEQUENCE 483 AA; 54215 MW; 1372ED3D CRC32;

Query Match 52.6%; Score 41; DB 1; Length 483;
 Best Local Similarity 58.3%; Pred. No. 32;
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 AITKLEGFDT 15
 I:|||||:
 Db 348 AVFKMELEGFET 359

RESULT 3
 ID 068607 PRELIMINARY; PRT; 355 AA.
 AC 068607;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE LYSYL-TRNA SYNTHETASE (FRAGMENT).
 GN LYS.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC11166;
 RA VAN VIET A.H.M., WOOLDRIDGE K.G., KETLEY J.M.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF052056; AAC64260.1; -.
 DR PFAM: PF00152; tRNA-synt_2; 1.
 KW Aminoacyl-tRNA synthetase.
 FT NON_TER 355 355
 SQ SEQUENCE 355 AA; 41238 MW; 140A75D0 CRC32;

Query Match 52.6%; Score 41; DB 2; Length 355;
 Best Local Similarity 53.8%; Pred. No. 23;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 EOAITKLEGFDT 14
 I:|||||:
 Db 338 EKILTKLADGFE 350

RESULT 4
 ID 064446 PRELIMINARY; PRT; 1462 AA.
 AC 064446;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE COPPER-TRANSPORTING P-TYPE ATPASE.
 GN ATP7B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DAT; TISSUE=LIVER;
 RX MEDLINE: 97049969
 RA THEOPHILOS M.B., COX D.W., MERCER J.F.B.;
 RT "The toxic milk mouse is a murine model of Wilson disease.";
 RL Hum. Mol. Genet. 5:1619-1624(1996).
 DR EMBL: U38477; AAC52852.1; -.
 DR MGD: MGI:103297; ATP7B.
 DR PROSITE: PS00154; ATPASE_E1_E2; 1.
 DR PFAM: PF00122; E1-E2_ATPase; 1.
 DR PFAM: PF00403; HMA; 5.
 KW Hydrolase; Transmembrane; Phosphorylation; ATP-binding.
 FT MOD_RES 1029 1029 PHOSPHORYLATION (PROBABLE).
 SQ SEQUENCE 1462 AA; 157296 MW; C53623A4 CRC32;

Query Match 52.6%; Score 41; DB 11; Length 1462;
 Best Local Similarity 43.8%; Pred. No. 1e+02;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEQAITKLEGFDTQ 16
 I:|||||:
 Db 150 AQEAVKLRVEGTCQ 165

RESULT 5
 ID 018258 PRELIMINARY; PRT; 489 AA.
 AC 018258;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
 DE Y7A9C.3 PROTEIN.
 GN Y7A9C.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MCMURRAY A.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: Z99286; CAB16543.1; -.
 SQ SEQUENCE 489 AA; 57598 MW; 86164B1F CRC32;

Query Match 51.3%; Score 40; DB 5; Length 489;
 Best Local Similarity 70.0%; Pred. No. 48;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 AITKLEGF 13
 I:|||||:

Db 426 SITKIKLQGF 435

RESULT 6

ID O46025 PRELIMINARY; PRT: 2561 AA.

AC O46025; 062379;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)

DE T12D8.1 PROTEIN.

GN T12D8.1

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

[1]

RN SEQUENCE FROM N.A.

RA STEWARD C.;

RN Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

[2]

RN SEQUENCE FROM N.A.

RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WORLDMAN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

[3]

RN SEQUENCE FROM N.A.

RA MCMURRAY A.;

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; 282094; CAB05024.1; -.

DR EMBL; 281120; CAB05024.1; JOINED.

DR EMBL; 281120; CAB03348.1; -.

DR EMBL; 282094; CAB03348.1; JOINED.

DR PFAM; PF00628; PHD; 1.

DR PFAM; PF00856; SET; 1.

SQ SEQUENCE 2561 AA; 286099 MW; 8E1BA84D CRC32;

Query Match 51.3%; Score 40; DB 5; Length 2561;

Best Local Similarity 62.5%; Pred. No. 2.7e+02;

Matches 10; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 2 EQAITKLK--KLEGFDT 15

|||||: | ||||

Db 2256 ENAITKITESLPGFDT 2271

RESULT 7

ID O53863 PRELIMINARY; PRT: 275 AA.

AC O53863;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)

DE PUTATIVE OXIDOREDUCTASE.

GN MV043.44C.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

[1]

RN SEQUENCE FROM N.A.

RA STRAIN-H37RV;

RA HAMLIN N., CHURCHER C.M.;

Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

[2]

RN SEQUENCE FROM N.A.

RC STRAIN-H37RV;

RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

[3]

RN SEQUENCE FROM N.A.

RP MEDLINE; 96181548.

RX PHILIPP W.J., POULET S., EICLMEIER K., PASCOPELLA L., JACOBS W.R. JR.,

RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R.,

RA COLE S.T.;

RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium

RT leprae.";

RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).

DR EMBL; AL022004; CAAL17657.1; -.

DR PFAM; PF00106; adh_short; 1.

SQ SEQUENCE 275 AA; 29249 MW; 594BEBB1 CRC32;

Query Match 50.0%; Score 39; DB 2; Length 275;

Best Local Similarity 58.3%; Pred. No. 39;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 QAITKLKLEGF 14

||: |: ||||

Db 45 QAVNHLRAEGFD 56

RESULT 8

ID O68274 PRELIMINARY; PRT: 505 AA.

AC O68274;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)

DE HYPOTHETICAL 55.1 KD PROTEIN.

OS Alcaligenes eutrophus.

OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;

OC Ralstonia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-H16;

RX MEDLINE; 98215161.

RA SLATER S., HOUMIEL K.L., TRAN M., MITSKY T.A., TAYLOR N.B.,

RA PADGETTE S., GRUYS K.J.;

RT "Multiple beta-ketothiolases mediate poly(beta-hydroxyalkanoate)

RT copolymer synthesis in Ralstonia eutropha.";

RL J. Bacteriol. 180:1979-1987(1998).

RN [2]

RP REVISIONS.

RA SLATER S.C.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF026544; AAC38321.2; -.

DR PFAM; PF00919; UPF0004; 1.

DR PROSITE; PS01278; UPF0004; 1.

KW Hypothetical protein.

SQ SEQUENCE 505 AA; 55100 MW; 47A1B157 CRC32;

Query Match 50.0%; Score 39; DB 2; Length 505;

Best Local Similarity 53.8%; Pred. No. 73;

Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEQAITKLKLEGF 13

||| |||: ||:

Db 77 SEQIIITLRAEGY 89

RESULT 9

Q24302

ID Q24302 PRELIMINARY; PRT; 934 AA.
AC Q24302;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)
DE CGMP-DEPENDENT PROTEIN KINASE (DG2;CD5).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89278147.
RA KALDERON D., RUBIN G.M.;
RT "CGMP-dependent protein kinase genes in Drosophila.";
RL J. Biol. Chem. 264:10738-10748(1989).
DR EMBL; M30149; AAA28454.1; .
DR FLYBASE; FBgn0000721; for.
DR PFAM; PF00027; CNMP_binding; 2.
DR PFAM; PF00069; pkinase; 1.
DR PFAM; PF00433; pkinase_C; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00888; CNMP_BINDING_1; 1.
DR PROSITE; PS00889; CNMP_BINDING_2; 2.
KW Kinase.
SQ SEQUENCE 934 AA; 105959 MW; B6417203 CRC32;

Query Match 50.0%; Score 39; DB 5; Length 934;
Best Local Similarity 81.8%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EQAITKLEK 12
|||||
Db 291 EQAIEGLKLEG 301

RESULT 10
Q24303
ID Q24303 PRELIMINARY; PRT; 894 AA.
AC Q24303;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)
DE CGMP-DEPENDENT PROTEIN KINASE (DG2;T2).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89278147.
RA KALDERON D., RUBIN G.M.;
RT "CGMP-dependent protein kinase genes in Drosophila.";
RL J. Biol. Chem. 264:10738-10748(1989).
DR EMBL; M27124; AAA28457.1; .
DR EMBL; M27122; AAA28457.1; JOINED.
DR EMBL; M27123; AAA28457.1; JOINED.
DR FLYBASE; FBgn0000721; for.
DR PFAM; PF00027; CNMP_binding; 2.
DR PFAM; PF00069; pkinase; 1.
DR PFAM; PF00433; pkinase_C; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00888; CNMP_BINDING_1; 1.
DR PROSITE; PS00889; CNMP_BINDING_2; 2.
KW Kinase.
SQ SEQUENCE 894 AA; 101025 MW; 6D4A143D CRC32;

Query Match 50.0%; Score 39; DB 5; Length 894;
Best Local Similarity 81.8%; Pred. No. 1.3e+02;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EQAITKLEK 12
|||||
Db 291 EQAIEGLKLEG 301

RESULT 11
O97961
ID O97961 PRELIMINARY; PRT; 1330 AA.
AC O97961;
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMELrel. 10, Last annotation update)
DE KINECTIN.
OS Vulpes vulpes (Red fox).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Vulpes.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-TESTIS;
RA XU J., JANSSENS P., BIRD P., EAST P., BRADLEY M.P.;
RT "Identification and characterization of fox kinectin.";
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF095786; AAC64407.1; .
SQ SEQUENCE 1330 AA; 153477 MW; 0D52375D CRC32;

Query Match 50.0%; Score 39; DB 6; Length 1330;
Best Local Similarity 63.6%; Pred. No. 2e+02;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEQAITKLEK 11
|||||
Db 479 AEQAVTQKQV 489

RESULT 12
Q48977
ID Q48977 PRELIMINARY; PRT; 119 AA.
AC Q48977;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)
DE Cl-THF SYNTHASE (FRAGMENT).
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC capricolum group.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 27343(KID);
RX MEDLINE; 96059641.
RA BORK P., OUZOUNIS C., CASARI G., SCHNEIDER R., SANDER C., DOLAN M.,
RA GILBERT W., GILLET P.M.;
RT "Exploring the Mycoplasma capricolum genome: a minimal cell reveals
its physiology.";
RL Mol. Microbiol. 16:955-967(1995).
DR EMBL; Z33044; CAAB3718.1; .
FT NON_TER 119
SQ SEQUENCE 119 AA; 13226 MW; 1A7E1F2A CRC32;

Query Match 49.4%; Score 38.5; DB 2; Length 119;
Best Local Similarity 42.3%; Pred. No. 20;
Matches 11; Conservative 1; Mismatches 1; Indels 13; Gaps 1;

Qy 3 QAITKLEK-----LEGFD 15
|||||
Db 42 QAITPLKVDGFHYINQKLEGYDT 67

RESULT 13
O29571
ID O29571 PRELIMINARY; PRT; 222 AA.
AC O29571;

DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-AUG-1998 (TRENBLrel. 07, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN AF0686.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE: 98049343.
 RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,
 RA FLEISCHMAN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 RA PETERSON S., REICH C.J., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
 RA OVERBEK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.;
 RT "The complete genome sequence of the hyperthermophilic,
 RT sulphate-reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:384-370(1997).
 DR EMBL; AE001057; AAB90552.1; -.
 DR TIGR; AF0686; -.
 KW Hypothetical protein.
 SQ SEQUENCE 222 AA; 25943 MW; 34044B45 CRC32;

Query Match 48.7%; Score 38; DB 1; Length 222;
 Best Local Similarity 57.1%; Pred. NO. 46;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 AEQAITKIKLEGFD 14
 ||| | : | | | |
 Db 85 AEQMISSKFEGMD 98

RESULT 14
 P77641
 ID P77641 PRELIMINARY; PRT; 296 AA.
 AC P77641;
 DT 01-FEB-1997 (TRENBLrel. 02, Created)
 DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
 DE FROM BASES 576335 TO 592504 (SECTION 51 OF 400) OF THE COMPLETE GENOME
 DE (SECTION 51 OF 400).
 GN YBCH.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K-12;
 RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T.,
 RA GLASNER F.D.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA CHUNG E., ALLEN E., ARAUJO R., APARICIO A., DAVIS K., DUNCAN M.,
 RA FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O., LEW H., LIN D.,
 RA NAMATH A., OEFNER P., ROBERTS D., SCHRAMM S., DAVIS R.W.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 97061202.
 RA OSHIMA T., AJBA H., BABA T., FUJITA K., HAYASHI K., HONTO A.,
 RA IKEMOTO K., INADA T., ITOH T., KAJIHARA M., KANAI K., KASHIMOTO K.,
 RA KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,

RA MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SAITO N.,
 RA SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,
 RA YANO M., HORIUCHI T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 DR EMBL; AE000161; AAC73868.1; -.
 DR EMBL; U82598; AAB40765.1; -.
 DR EMBL; D90699; BAA35201.1; -.
 SQ SEQUENCE 296 AA; 33056 MW; C24CA8B6 CRC32;

Query Match 48.7%; Score 38; DB 2; Length 296;
 Best Local Similarity 54.5%; Pred. NO. 62;
 Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 5 ITKIKLEGFD 15
 ::||:|:||||
 Db 44 MSQRLQGFDT 54

RESULT 15
 Q9ZKP8
 ID Q9ZKP8 PRELIMINARY; PRT; 188 AA.
 AC Q9ZKP8;
 DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
 DE PUTATIVE.
 GN JHP0887.
 OS Helicobacter pylori J99.
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-J99;
 RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
 RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
 RA TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
 RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
 RA TRUST T.J.;
 RT "Genomic Sequence Comparison of Two Unrelated Isolates of the Human
 RT Gastric Pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 DR EMBL; AE001518; AAD06471.1; -.
 SQ SEQUENCE 188 AA; 21317 MW; AC1E5216 CRC32;

Query Match 48.7%; Score 38; DB 2; Length 188;
 Best Local Similarity 58.3%; Pred. NO. 39;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEQAITKIKLEG 12
 :| | :|:||||
 Db 137 SEVAVVKIKLEG 148

Search completed: November 13, 1999, 12:55:26
 Job time: 3025 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:59:48 ; Search time 104.22 Seconds
(without alignments)
3.182 Million cell updates/sec

Title: US-08-913-430-6
Perfect score: 71
Sequence: 1 KNSQNKIIDLSPEG 14

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	71	100.0	132	1 R06277	Swine enzootic pne
2	71	100.0	120	1 R06278	Swine enzootic pne
3	71	100.0	281	1 R06279	Swine enzootic pne
4	71	100.0	419	1 R21829	Sequence of surfac
5	71	100.0	14	1 W01036	Mycoplasma 46-48 k
6	71	100.0	419	1 W01037	Mycoplasma 46-48 k
7	45	63.4	398	1 W11798	Dirofilaria immiti
8	45	63.4	314	1 W11801	Dirofilaria immiti
9	45	63.4	400	1 W69545	Dirofilaria immiti
10	45	63.4	316	1 W69546	Dirofilaria immiti
11	38	53.5	329	1 W23388	Shaker-like potass
12	37	52.1	878	1 W20966	H. pylori flagella
13	37	52.1	394	1 W20578	H. pylori flagella
14	37	52.1	394	1 W24694	H. pylori flagella
15	37	52.1	731	1 W44849	S. pneumoniae peni
16	36	50.7	387	1 P90400	Modified human lip
17	36	50.7	346	1 P82063	Human lipocortin R
18	36	50.7	346	1 R06560	Human lipocortin O
19	36	50.7	363	1 P60657	Sequence of human
20	36	50.7	346	1 P61523	Sequence of human
21	36	50.7	363	1 R22402	Human lipocortin.
22	36	50.7	452	1 R97299	Human GABA-A recep
23	36	50.7	453	1 W24250	Aquifex adenosyl-8
24	36	50.7	264	1 W70475	Girdwood S. A. virus
25	36	50.7	264	1 W70482	Sindbis virus caps
26	36	50.7	284	1 W70468	South African Arbo
27	35	49.3	418	1 P94664	Predominant form o
28	35	49.3	399	1 R04033	GAPDH promotor fra
29	35	49.3	394	1 R03754	Entire sequence of
30	35	49.3	308	1 R13073	Serine protease in
31	35	49.3	394	1 P61708	[Gly358] alphas-an
32	35	49.3	394	1 P61709	[Ala358] alphas-an
33	35	49.3	394	1 P61710	[Ile358] alphas-an
34	35	49.3	394	1 P61711	[Ile358] alphas-an
35	35	49.3	394	1 P61712	[Leu358] alphas-an
36	35	49.3	394	1 P61713	[Phe358] alphas-an
37	35	49.3	394	1 P60512	[Arg358] alphas-an
38	35	49.3	418	1 P50877	Sequence encoded b
39	35	49.3	418	1 P50021	Sequence of alpha-
40	35	49.3	418	1 P90128	Sequence encoded b
41	35	49.3	393	1 R20802	Alpha-1-antitrypsi
42	35	49.3	418	1 P40133	Sequence of human
43	35	49.3	418	1 R22931	Alpha-1 antitrypsi

Recombinant M24-M5
Human alpha-1-anti

44 35 49.3 305 1 R50996
45 35 49.3 394 1 R67360

ALIGNMENTS

RESULT 1
R06277 ID R06277 standard; protein; 132 AA.
AC R06277;
DT 19-DEC-1990 (first entry)
DE Swine enzootic pneumonia mycoplasma surface antigen polypeptide.
KW Mycoplasma pneumoniae; enzootic pneumonia; ds.
OS Mycoplasma hyopneumoniae.
PN J02167079-A.
PD 27-JUN-1990.
PF 21-DEC-1988; 322829.
PR 21-DEC-1988; JP-322829.
PA (NISE-) NIPPON SEIFUN KK.
DR WPI; 90-241949/32.
DR N-PSDB; Q05576.
PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used for early detection and treatment of mycoplasma or enzootic pneumonia of pigs
PS Claim 1; Page 570; 28pp; Japanese.
CC Sequence encoding polypeptides may be used as probes for the early diagnosis of pneumonia in pigs. The sequences may be used to produce an expression vector and transform a suitable host.
SQ Sequence 132 AA;

Query Match 100.0%; Score 71; DB 1; Length 132;
Best Local Similarity 100.0%; Pred. No. 9.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPEG 14
DB 74 KNSQNKIIDLSPEG 87

RESULT 2
R06278 ID R06278 standard; protein; 120 AA.
AC R06278;
DT 19-DEC-1990 (first entry)
DE Swine enzootic pneumonia mycoplasma surface antigen polypeptide.
KW Mycoplasma pneumoniae; enzootic pneumonia; ds.
OS Mycoplasma hyopneumoniae.
PN J02167079-A.
PD 27-JUN-1990.
PF 21-DEC-1988; 322829.
PR 21-DEC-1988; JP-322829.
PA (NISE-) NIPPON SEIFUN KK.
DR WPI; 90-241949/32.
DR N-PSDB; Q05577.
PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used for early detection and treatment of mycoplasma or enzootic pneumonia of pigs
PS Claim 1; Page 570; 28pp; Japanese.
CC Sequence encoding polypeptides may be used as probes for the early diagnosis of pneumonia in pigs. The sequences may be used to produce an expression vector and transform a suitable host.
SQ Sequence 120 AA;

Query Match 100.0%; Score 71; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 8.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPEG 14
DB 80 KNSQNKIIDLSPEG 93

RESULT 3
 R06279 ID R06279 standard; protein; 281 AA.
 AC R06279; 19-DEC-1990 (first entry)
 DT DE Swine enzootic pneumonia mycoplasma surface antigen polypeptide.
 DE Swine enzootic pneumonia; enzootic pneumonia; ds.
 KW Mycoplasma hyopneumoniae.
 OS Mycoplasma hyopneumoniae.
 PN J02167079-A.
 PD 27-JUN-1990.
 PF 21-DEC-1988; 322829.
 PR 21-DEC-1988; JP-322829.
 PA (NISE-) NIPPON SEIFUN KK.
 DR WPI: 90-241949/32.
 DR N-PSDB; Q05378.
 PT Surface antigen DNA of swine enzootic pneumonia mycoplasma - used for early detection and treatment of mycoplasma or enzootic pneumonia of pigs
 PT Claim 1; page 570; 28pp; Japanese.
 PS Sequence encoding polypeptides may be used as probes for the early diagnosis of pneumonia in pigs. The sequences may be used to produce an expression vector and transform a suitable host.
 CC
 SQ Sequence 281 AA;

Query Match 100.0%; Score 71; DB 1; Length 281;
 Best Local Similarity 100.0%; Pred. No. 2.1e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPEG 14
 |||||
 DB 94 KNSQNKIIDLSPEG 107

RESULT 4
 R21829 ID R21829 standard; Protein; 419 AA.
 AC R21829;
 DT 02-NOV-1992 (first entry)
 DE Sequence of surface antigen 46kd.
 KW Swine pneumonia; epidemic; diagnosis; therapy.
 OS Mycoplasma hyopneumoniae.
 PN EP-475185-A.
 PD 18-MAR-1992.
 PF 27-AUG-1991; 114335.
 PR 27-AUG-1990; JP-224945.
 PA (NIFL-) NIPPON FLOUR MILLS.
 PI Seto Y, Futo S, Mitsuse S, Matsuo K, Tsuna M;
 DR WPI: 92-089874/12.
 DR N-PSDB; Q22042.
 PT DNA and peptide of mycoplasma hyopneumoniae - useful for diagnosis and treatment of swine mycoplasma pneumonia
 PS Disclosure; Page 4-6 and pages 34-36; 45pp; English.
 CC The inventors claim DNA encoding a surface antigen and primers used in a method of diagnosing mycoplasma pneumonia of swine.
 CC Mycoplasma cells were collected from the culture of M.hp. lysed in a buffer containing SDS, followed by the extraction of DNA and purification thereof. The DNA of M.hp. is cleaved with a restriction enzyme HindIII, the resulting fragment is inserted into a plasmid pUC119 at the site cleaved with HindIII. This plasmid is transformed into E. coli as a host cell, colony hybridization is performed by the DNA probe method utilizing EcoRI fragments of pKUM1 and pKUM2 to select the bacterial cells containing a plasmid (pUR126) carrying the 46 kd antigenic gene. E. coli JM 109 which comprises the plasmid pKUM1 or pKUM2 is deposited with FRI under accession No. FERM P-10318 or P-10319. The base sequence of the 46 kd antigen gene is given in Q22042.
 CC
 SQ Sequence 419 AA;

Query Match 100.0%; Score 71; DB 1; Length 419;

Best Local Similarity 100.0%; Pred. No. 3.3e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPEG 14
 |||||
 DB 232 KNSQNKIIDLSPEG 245

RESULT 5
 W01036 ID W01036 standard; Peptide; 14 AA.
 AC W01036;
 DT 19-JAN-1997 (first entry)
 DE Mycoplasma 46-48 kDa protective antigen fragment CNBR F3.
 KW Antigen; vaccine; mycoplasma pneumonia; swine enzootic pneumonia; diagnosis; antibody.
 KW
 OS Mycoplasma hyopneumoniae strain Beaufort.
 PN W09628472-A1.
 PD 19-SEP-1996.
 PF 15-MAR-1996; AU0149.
 PR 16-MAR-1995; AU-001789.
 PA (UYME) UNIV MELBOURNE.
 PI Doughty SW, Lee R, Walker J;
 DR WPI: 96-433763/43.
 PT Putative protective antigens against Mycoplasma - used for the detection, prevention or treatment of Mycoplasma infections, esp. M. hyopneumoniae in swine
 PT Claim 13; Page 28; 43pp; English.
 PS A 46-48 kDa putative protective antigen against Mycoplasma contains the N-terminal sequence given in W01033 and the internal CNBR fragments given in W01034-36. The antigen was isolated from Mycoplasma hyopneumoniae cells using antibody probes enriched with Mycoplasma-specific antibodies. Other protective antigens were also identified (see also W01024-32). A gene (T38241) coding for the 48 kDa antigen (W01037) was isolated from a genomic library. Protective antigens and antibodies can be used in vaccines for preventing or treating mycoplasma infections, partic. M. hyopneumoniae infections in swine. They can also be used for diagnosis.
 SQ Sequence 14 AA;

Query Match 100.0%; Score 71; DB 1; Length 14;
 Best Local Similarity 100.0%; Pred. No. 7.7e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPEG 14
 |||||
 DB 1 KNSQNKIIDLSPEG 14

RESULT 6
 W01037 ID W01037 standard; Protein; 419 AA.
 AC W01037;
 DT 19-JAN-1997 (first entry)
 DE Mycoplasma 46-48 kDa protective antigen.
 KW Antigen; vaccine; mycoplasma pneumonia; swine enzootic pneumonia; diagnosis; antibody.
 KW
 OS Mycoplasma hyopneumoniae strain Beaufort.
 FH Key Location/Qualifiers
 FT misc_difference 70
 FT /note= "codon 70 in the nucleotide sequence is a stop codon"
 FT
 FT misc_difference 101
 FT /note= "codon 101 in the nucleotide sequence is a stop codon"
 FT
 FT misc_difference 254
 FT /note= "codon 254 in the nucleotide sequence is a stop codon"
 FT
 FT
 PN W09628472-A1.
 PD 19-SEP-1996.
 PF 15-MAR-1996; AU0149.

DR N-PSDB; V40248.
PT Nematode larval protease proteins - useful for vaccination, etc.
PS Example 1; Column 31-34; 22pp; English.
CC The present sequence represents an L3 larval protease protein from
CC Dirofilaria immitis. An embodiment of the present invention is an
CC isolated filarid nematode nucleic acid molecule that hybridises, under
CC stringent hybridisation conditions, with a Dirofilaria immitis L3 larval
CC cysteine protease gene and/or an Onchocerca volvulus L3 larval cysteine
CC protease gene. A filarid nematode cysteine protease protein of the
CC present invention preferably has cysteine protease activity and/or
CC comprises a protein that, when administered to an animal, is capable of
CC eliciting an immune response against a natural helminth cysteine protease
CC protein. This sequence can be used in a therapeutic composition
CC capable of protecting an animal from disease caused by a parasitic
CC helminth.
SQ Sequence 400 AA;

Query Match 63.4%; Score 45; DB 1; Length 400;
Best Local Similarity 53.8%; Pred. No. 1.7;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPE 13
| : : : : :
DB 224 KTKNKLLDLSFQ 236

RESULT 10

W69346
ID W69546 standard; Protein; 316 AA.

AC W69546;
DT 09-OCT-1998 (first entry)
DE Dirofilaria immitis L3 larval cysteine protease #2.
KW Onchocerca volvulus; Dirofilaria immitis; L3 larval cysteine protease;
KW filarid nematode; parasitic helminth; vaccine.
OS Dirofilaria immitis.
PN US5792624-A.
PD 11-AUG-1998.
PF 07-JUN-1995; 482282.
PR 07-JUN-1995; US-482282.
PR 12-FEB-1991; US-654226.
PR 12-NOV-1991; US-792209.
PR 03-AUG-1993; US-101283.
PR 16-NOV-1993; US-153554.
PA (HESK-) HESKA CORP.
PA (COLS-) UNIV COLORADO STATE RES FOUND.

PI Frank GR, Griewe RB, Richer JK, Tripp CA, Wisniewski N;
DR WPI; 98-456128/39.
DR N-PSDB; V40249.
PT Nematode larval protease proteins - useful for vaccination, etc.
PS Example 1; Column 35-38; 22pp; English.
CC The present sequence represents an L3 larval protease protein from
CC Dirofilaria immitis. An embodiment of the present invention is an
CC isolated filarid nematode nucleic acid molecule that hybridises, under
CC stringent hybridisation conditions, with a Dirofilaria immitis L3 larval
CC cysteine protease gene and/or an Onchocerca volvulus L3 larval cysteine
CC protease gene. A filarid nematode cysteine protease protein of the
CC present invention preferably has cysteine protease activity and/or
CC comprises a protein that, when administered to an animal, is capable of
CC eliciting an immune response against a natural helminth cysteine protease
CC protein. This sequence can be used in a therapeutic composition
CC capable of protecting an animal from disease caused by a parasitic
CC helminth.
SQ Sequence 316 AA;

Query Match 63.4%; Score 45; DB 1; Length 316;
Best Local Similarity 53.8%; Pred. No. 1.3;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPE 13
| : : : : :
DB 140 KTKNKLLDLSFQ 152

RESULT 11

W23388
ID W23388 standard; Protein; 329 AA.

AC W23388;
DT 02-APR-1998 (first entry)
DE Shaker-like potassium ion channel beta-subunit core region Kv beta 1.
DE Shaker-like potassium ion channel; SPC; beta subunit; core region;
KW Shaker-like potassium ion channel; SPC; beta subunit; core region;
KW Kv beta 1; N-terminal A and B box; NAB; treatment; cardiac disease;
KW tumour; auto immune disease.
OS Mammalia.

PN WO9731112-A2.
PD 28-AUG-1997.
PF 18-FEB-1997; U02292.
PR 23-FEB-1996; US-606143.
PA (UJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PI Li M;

DR WPI; 97-435164/40.

PT Polypeptide(s) derived from Shaker-like potassium ion channel alpha
and beta subunits - used to alter potassium ion levels in a cell.
PT e.g. for treating neurological disorders, tumours, metabolic disease
PT and cardiac disease

PS Claim 15; Pages 51-52; 106pp; English.

CC This polypeptide sequence Kv beta 1 consists of the core region of a
beta-subunit of a Shaker-like potassium ion channel (SPC). This core
region can bind to a polypeptide consisting of the N-terminal A and B
box (NAB) domain and the NAB-S1 (the first transmembrane spanning domain)
linking region of the alpha-subunit of SPC. The polypeptides or the
nucleic acid encoding them can be introduced into the cytoplasm of a
cell to modulate the flow of potassium ions through a cytoplasmic cell
membrane. Potassium ion channels regulate the action potentials, cardiac
pacemaking and neurotransmitter release in excitable tissues. In
non-excitabile tissues they play important roles in hormone secretion,
cell proliferation, cell volume regulation and lymphocyte
differentiation. Molecules which bind to the alpha or beta-subunit
polypeptides can be detected by contacting the polypeptides with a
putative NAB and NAB-S1 linking region of an alpha-subunit or with a
putative core region of a beta-subunit respectively, and determining
whether or not binding occurs. These polypeptides and the encoding
nucleic acids may be useful in the treatment of a huge variety of
disorders, e.g. neurological disorders, tumours, metabolic diseases,
cardiac disease and autoimmune disease.

SQ Sequence 329 AA;

Query Match 53.5%; Score 38; DB 1; Length 329;
Best Local Similarity 58.3%; Pred. No. 25;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSP 12
| : : : : :
DB 245 RKQONKLLDLSP 256

RESULT 12

W20966
ID W20966 standard; Protein; 878 AA.

AC W20966;
DT 21-JUL-1997 (first entry)
DE H. pylori flagella-associated protein, hp2el091lorf5.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.

OS Helicobacter pylori.

PN WO9640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; U09122.

PR 07-JUN-1995; US-487032.

PR 01-APR-1996; US-630405.

PA (ASTR) ASTRA AB.

PI Berglinth OT, Smith D, Mellgaard BL;

DR WPI: 97-052306/05.
 DR N-PSDB; T68219.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 56; Page 1356-57; 1481pp; English.
 CC This sequence represents a H. pylori flagella-associated protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 878 AA;

Query Match 52.1%; Score 37; DB 1; Length 878;
 Best Local Similarity 66.7%; Pred. No. 1; Indels 0; Gaps 0;
 Matches 6; Conservative 3; Mismatches 0;

QY 3 SONKIIDL 11
 |||||:|:
 Db 341 SQNKVDVS 349

RESULT 13
 W20578
 ID W20578 standard; Protein; 394 AA.
 AC W20578; 1997 (first entry)
 DT 04-AUG-1997
 DE H. pylori flagella-associated protein 917152 aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis.
 OS Helicobacter pylori.
 FH Key Location/Qualifiers
 FT misc_difference 140
 FT /label= "unknown"
 FT /note= "encoded by CAM"

W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OF, Smith D, Mellgaerd BL;
 DR WPI: 97-052306/05.
 DR N-PSDB; T67832.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 56; Page 731; 1481pp; English.
 CC This sequence shows a Helicobacter pylori flagella-associated
 CC protein that may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.
 SQ Sequence 394 AA;

Query Match 52.1%; Score 37; DB 1; Length 394;
 Best Local Similarity 66.7%; Pred. No. 47;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONKIIDL 11
 |||||:|:
 Db 229 SQNKVDVS 237

RESULT 14
 W24694
 ID W24694 standard; Protein; 394 AA.
 AC W24694; 1997 (first entry)
 DT 13-AUG-1997
 DE H. pylori flagella-associated protein 867183 aa.
 KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
 KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
 KW detection; antisense; inhibition.
 OS Helicobacter pylori.
 FH Key Location/Qualifiers
 FT misc_difference 418.420
 FT /note= "encoded by CAM"

W09719098-A1.
 PD 29-MAY-1997.
 PF 15-NOV-1996; U18542.
 PR 17-NOV-1995; US-561469.
 PA (ASTR) ASTRA AB.
 PI Smith DH;
 DR WPI: 97-298052/27.
 DR N-PSDB; T77512.
 PT Helicobacter pylori nucleic acid sequences and related proteins -
 PT used for diagnostics and therapeutics
 PS Claim 18; Page 198; 235pp; English.
 CC This sequence shows an H. pylori flagella-associated protein that
 CC shows homology to minor flagellin flab precursor protein.
 CC Helicobacter pylori has been strongly linked to chronic gastritis and
 CC duodenal ulcer disease. The nucleic acid sequences of the invention
 CC are used to evaluate compounds, especially activators or inhibitors of
 CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid
 CC sequence. The nucleic acid sequences, and corresponding proteins, are
 CC also useful for generating vaccines for immunising subjects against H.
 CC pylori or for use in detecting the presence of Helicobacter species in
 CC a sample. Antisense nucleic acid sequences of these sequences are
 CC used to inhibit expression of a gene from Helicobacter species. H.
 CC pylori whole genomic DNA was isolated and nebulised to a median size of
 CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
 CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
 CC complementary to the BstXI-cut PMPX vectors, while the overhang is not
 CC self-complementary. Therefore the linkers will not concatamerise nor
 CC will the cut vector re-ligate itself easily. The linker-adaptor inserts
 CC were ligated to each of the 20 PMPX vectors to construct a series of
 CC shotgun subclone libraries. The purified DNA samples were then
 CC sequenced.
 CC Note: the ORF/protein reference number for this sequence was obtained
 CC from the related specification, W09640893.
 SQ Sequence 394 AA;

Query Match 52.1%; Score 37; DB 1; Length 394;
 Best Local Similarity 66.7%; Pred. No. 47;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 SONKIIDL 11
 |||||:|:
 Db 229 SQNKVDVS 237

RESULT 15

W44849
ID W44849 standard; Protein; 731 AA.
AC W44849;
DE 21-AUG-1998 (first entry)
DE S. pneumoniae penicillin-binding protein PBP-Nv.
KW Penicillin-binding protein; PBP-Nv; transglycolase; antibacterial agent;
KW trans-peptidation activity; infection; therapy.
OS Streptococcus pneumoniae.
PN EP-837132-A2.
PN 22-APR-1998.
PD 17-OCT-1997; 308288.
PF 17-OCT-1996; US-731716.
PR (ELIL) LILLY & CO ELI.
PI Hoskins JA, Jaskunas SR, Norris FH, Rockey PK, Rosteck PR,
PI Zhao G;
DR WPI: 98-219114/20.
DR N-PSDB; V19373.
DR Recombinant Streptococcus pneumoniae penicillin-binding protein -
PT useful in screening assays for antibacterial agents
PS Claim 1; Page 19-21; 28pp; English.
CC This sequence is the Streptococcus pneumoniae penicillin-binding
CC protein, designated PBP-Nv, of the invention. The protein is useful in
CC screening assays for compounds that bind to PBP-Nv or inhibit the
CC transglycolase or trans-peptidation activity of PBP-Nv. Such compounds
CC would be useful as antibacterial agents for treating S. pneumoniae
CC infections.
SQ Sequence 731 AA;

Query Match 52.1%; Score 37; DB 1; Length 731;
Best Local Similarity 63.6%; Pred. No. 93;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KNSONKIIDL 11
|||||:||||
Db 549 KNSQRRVIDKS 559

Search completed: November 13, 1999, 18:59:49
Job time: 128 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:29 ; Search time 64.87 Seconds
(without alignments)
2.466 Million cell updates/sec

Title: US-08-913-430-6
Perfect score: 71
Sequence: 1 KNSQNKIIDLSPG 14
Scoring table: BLOSUM62
Searched: 119832 seqs, 11428610 residues
Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/PCTUS9.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	63.4	400	2	US-08-482-282B-2
2	45	63.4	316	2	US-08-482-282B-4
3	45	63.4	398	2	US-08-486-036A-2
4	45	63.4	314	3	US-08-486-036A-4
5	45	63.4	398	3	PCT-US96-09848-2
6	45	63.4	314	3	PCT-US96-09848-4
7	45	63.4	213	3	PCT-US96-09848-21
8	45	63.4	356	3	PCT-US96-09848-24
9	38	53.5	329	2	US-08-606-143-1
10	38	53.5	329	2	US-08-606-143-3
11	37	52.1	731	2	US-08-731-716-2
12	37	52.1	102	2	US-08-710-749-21
13	36	50.7	1245	2	US-08-801-263A-3
14	36	50.7	1245	2	US-08-801-263A-6
15	36	50.7	1245	2	US-08-801-263A-10
16	36	50.7	453	2	US-08-599-171A-27
17	36	50.7	386	2	US-08-559-303B-75
18	35	49.3	308	1	US-07-859-480-2
19	35	49.3	418	1	US-08-121-714-3
20	35	49.3	394	1	US-08-002-202-6
21	35	49.3	394	1	US-08-002-202-9
22	35	49.3	394	1	US-08-002-202-11
23	35	49.3	414	1	US-08-002-202-13
24	35	49.3	414	1	US-08-002-202-17
25	35	49.3	414	1	US-08-002-202-19
26	35	49.3	1169	1	US-08-542-921-2
27	35	49.3	418	2	US-08-477-108A-3
28	35	49.3	394	2	US-08-553-488A-1
29	35	49.3	1169	2	US-08-880-685-2
30	35	49.3	1169	2	US-08-880-684-2
31	35	49.3	2710	2	US-08-588-459A-12
32	35	49.3	528	2	US-08-793-229-35
33	35	49.3	418	2	US-08-477-112-3
34	35	49.3	418	3	PCT-US93-08322-3
35	34	47.9	178	1	US-08-370-975B-14
36	34	47.9	130	2	US-08-853-659A-39
37	34	47.9	488	2	US-08-933-750C-17
38	33	46.5	327	1	US-08-238-163-4
39	33	46.5	1235	1	US-08-118-101A-2

40 33 46.5 273 1 US-08-221-750A-9
41 33 46.5 557 1 US-08-328-256-10
42 33 46.5 434 1 US-08-328-256-11
43 33 46.5 496 1 US-08-328-256-12
44 33 46.5 2414 1 US-08-227-536-2
45 33 46.5 557 1 US-08-471-454-2

ALIGNMENTS

RESULT 1
US-08-482-282B-2
; Sequence 2, Application US/08482282B
; Patent No. 5792624
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80525

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect for Windows, Version 7.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,282B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: 2618-33-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (970) 493-7272
TELEFAX: (970) 484-9505
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-482-282B-2

Query Match 63.4%; Score 45; DB 2; Length 400;
Best Local Similarity 53.8%; Pred. No. 0.66;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPG 13
Db 224 KTKNKLIDLSPPQ 236

RESULT 2
US-08-482-282B-4
; Sequence 4, Application US/08482282B
; Patent No. 5792624
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy

```
;
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARILID NEMATODE CYSTEINE
; TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,282B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: 2618-33-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (970) 493-7272
; TELEFAX: (970) 484-9505
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-482-282B-4

Query Match 63.48; Score 45; DB 2; Length 316;
Best Local Similarity 53.88; Pred. No. 0.5;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSPE 13
Db 140 KTKNKLDSLSPQ 152

RESULT 3
US-08-486-036A-2
; Sequence 2, Application US/08486036A
; Patent No. 5795768
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARILID NEMATODE
; TITLE OF INVENTION: CYSTEINE PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,036A
; FILING DATE: June 7, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-486-036A-4
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,036A
; FILING DATE: June 7, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-486-036A-2

Query Match 63.48; Score 45; DB 2; Length 398;
Best Local Similarity 53.88; Pred. No. 0.65;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSPE 13
Db 224 KTKNKLDSLSPQ 236

RESULT 4
US-08-486-036A-4
; Sequence 4, Application US/08486036A
; Patent No. 5795768
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARILID NEMATODE
; TITLE OF INVENTION: CYSTEINE PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,036A
; FILING DATE: June 7, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-486-036A-4
```


REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-33-PCT
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-09848-21

Query Match 63.4%; Score 45; DB 3; Length 213;
Best Local Similarity 53.8%; Pred. No. 0.31;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSP 13
| : : : : :
Db 39 KTKNKLKLDLSPQ 51

RESULT 8
PCT-US96-09848-24
; Sequence 24, Application PC/TUS9609848
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09848
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,036
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1172
PCT-US96-09848-24

Query Match 63.4%; Score 45; DB 3; Length 356;
Best Local Similarity 53.8%; Pred. No. 0.57;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSP 13
| : : : : :
Db 182 KTKNKLKLDLSPQ 194

RESULT 9
US-08-606-143-1
; Sequence 1, Application US/08606143
; Patent No. 5856155
; GENERAL INFORMATION:
; APPLICANT: Li, Min
; TITLE OF INVENTION:
; TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH
; TITLE OF INVENTION: COMPOUNDS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60601

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/606,143
; FILING DATE: 23-FEB-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kilyk Jr., John
; REGISTRATION NUMBER: 30763
; REFERENCE/DOCKET NUMBER: 71756
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 329 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-606-143-1

Query Match 53.5%; Score 38; DB 2; Length 329;
Best Local Similarity 58.3%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSP 12
| : : : : :
Db 245 RKQNKLDLSP 256

RESULT 10
US-08-606-143-3
; Sequence 3, Application US/08606143
; Patent No. 5856155
; GENERAL INFORMATION:
; APPLICANT: Li, Min
; TITLE OF INVENTION:
; TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH
; TITLE OF INVENTION: COMPOUNDS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago

STATE: IL
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/606,143
FILING DATE: 23-FEB-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kilyk Jr., John
REGISTRATION NUMBER: 30763
REFERENCE/DOCKET NUMBER: 71756
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
TELEX: 25-3533
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-606-143-3

Query Match 53.5%; Score 38; DB 2; Length 329;
Best Local Similarity 58.3%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KNSONKIIDLSP 12
DB 245 KQONKLDLSP 256

RESULT 11
US-08-731-716-2
Sequence 2, Application US/08731716
Patent No. 5789202
GENERAL INFORMATION:
APPLICANT: Hoskins, JoAnn
APPLICANT: Jaskunas, S. Richard
APPLICANT: Rocky, Pamela K.
APPLICANT: Zhao, Genshi
APPLICANT: Rostock, Paul R. Jr.
APPLICANT: No. 5789202ris, Franklin H.
TITLE OF INVENTION: Penicillin Binding Protein From
TITLE OF INVENTION: Streptococcus Pneumoniae
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: U.S.
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/731,716
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Thomas D.
REGISTRATION NUMBER: 39,872
REFERENCE/DOCKET NUMBER: X-10,887

TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3334
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 731 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-731-716-2

Query Match 52.1%; Score 37; DB 2; Length 731;
Best Local Similarity 63.6%; Pred. No. 43;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNSONKIIDL 11
DB 549 KNSQKRVIDKS 559

RESULT 12
US-08-710-749-21
Sequence 21, Application US/08710749
Patent No. 5955089
GENERAL INFORMATION:
APPLICANT: Briles, David E.
APPLICANT: Hollingshead, Susan
APPLICANT: Becker, Robert
TITLE OF INVENTION: STRAIN SELECTION OF PNEUMOCOCCAL SURFACE
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/710,749
FILING DATE: 20-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454313-2074
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 amino acids
TYPE: amino acid
STRANDEDNESS: n/a
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-08-710-749-21

Query Match 52.1%; Score 37; DB 2; Length 102;
Best Local Similarity 66.7%; Pred. No. 4.2;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 KIIDLSP 14
DB 10 KLKLDLDP 18

RESULT 13
US-08-801-263A-3
; Sequence 3, Application US/08801263A
; Patent No. 5811407
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5811407th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/801,263A
; FILING DATE: 19-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-147
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1245 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-801-263A-3

Query Match 50.7%; Score 36; DB 2; Length 1245;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 NSQNKIIDLSPEG 14
||: | | : |||
Db 248 NSKGKTIKTPEG 260

RESULT 14
US-08-801-263A-6
; Sequence 6, Application US/08801263A
; Patent No. 5811407
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5811407th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,263A
; FILING DATE: 19-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-147
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1245 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-801-263A-6

Query Match 50.7%; Score 36; DB 2; Length 1245;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 NSQNKIIDLSPEG 14
||: | | : |||
Db 248 NSKGKTIKTPEG 260

RESULT 15
US-08-801-263A-10
; Sequence 10, Application US/08801263A
; Patent No. 5811407
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5811407th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,263A
; FILING DATE: 19-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-147
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1245 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-801-263A-10

Query Match 50.7%; Score 36; DB 2; Length 1245;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 2 NSQNKIIDLSPEG 14
||: ||: ||:
Db 248 NSKGKTIKTIPEG 260

Search completed: November 13, 1999, 10:56:29
Job time: 1358 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:34 ; Search time 251.81 Seconds
(without alignments)
3.520 Million cell updates/sec

Title: US-08-913-430-6

Perfect score: 71

Sequence: 1 KNSQNIIDLSPEG 14

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA.*
1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71	100.0	419	4	US-08-183-774-1
2	71	100.0	419	13	US-08-913-430-2
3	71	100.0	14	13	US-08-913-430-6
4	45	63.4	398	8	US-08-482-282A-2
5	45	63.4	314	8	US-08-482-282A-4
6	45	63.4	398	11	US-08-768-619-2
7	45	63.4	314	11	US-08-768-619-4
8	45	63.4	213	11	US-08-768-619-22
9	45	63.4	356	11	US-08-768-619-25
10	45	63.4	400	11	US-08-768-619-33
11	45	63.4	215	11	US-08-768-619-38
12	45	63.4	401	11	US-08-768-619-44
13	45	63.4	241	11	US-08-768-619-46
14	45	63.4	398	14	US-09-005-298-2
15	45	63.4	314	14	US-09-005-298-4
16	45	63.4	213	14	US-09-005-298-22
17	45	63.4	356	14	US-09-005-298-25
18	45	63.4	400	14	US-09-005-298-33
19	45	63.4	215	14	US-09-005-298-38

20	45	63.4	401	14	US-09-005-298-44
21	45	63.4	241	14	US-09-005-298-46
22	44	62.0	141	12	US-08-827-356-3336
23	44	62.0	156	16	US-09-276-872-2
24	40	56.3	499	16	US-09-248-796-14484
25	40	56.3	499	19	US-60-096-409-14484
26	39	54.9	327	16	US-09-248-796-17408
27	39	54.9	327	19	US-60-096-409-17408
28	38	53.5	210	15	US-09-107-532-4067
29	38	53.5	732	15	US-09-134-000-6359
30	38	53.5	363	16	US-09-248-796-19011
31	38	53.5	394	16	US-09-248-796-19011
32	38	53.5	452	17	US-09-328-352-7162
33	38	53.5	363	19	US-60-096-409-19011
34	38	53.5	394	19	US-60-096-409-19011
35	38	53.5	331	19	US-60-142-843-270
36	38	53.5	441	19	US-60-145-989-427
37	37	52.1	394	1	PCT-US96-18542-226
38	37	52.1	731	1	PCT-US97-19070-2
39	37	52.1	394	8	US-08-487-032A-877
40	37	52.1	394	9	US-08-561-469A-877
41	37	52.1	394	11	US-08-761-184-840
42	37	52.1	879	11	US-08-761-184-1262
43	37	52.1	828	11	US-08-761-184-1419
44	37	52.1	394	12	US-08-821-931-840
45	37	52.1	879	12	US-08-821-931-1262

ALIGNMENTS

RESULT 1
US-08-183-774-1
; Sequence 1, Application US/08183774
; GENERAL INFORMATION:
; APPLICANT: SETO, Yasuhiro
; APPLICANT: FOTO, Satoshi
; APPLICANT: MITSUSE, Shizuo
; APPLICANT: MATSUO, Kanako
; APPLICANT: TSUNA, Mika
; TITLE OF INVENTION: DNA's Encoding Surface Antigen of
; TITLE OF INVENTION: Mycoplasma hyopneumoniae, DNA Fragments for Primer,
; TITLE OF INVENTION: Recombinant Antigenic Peptides and Diagnostic Method of
; TITLE OF INVENTION: Mycoplasma Pneumoniae of Swine Using Same
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/183,774
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/747,015
; FILING DATE: 19-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, Norman F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 4183-001-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:

```

RESULT      3
US-08-913-430-6
; Sequence 6, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 6
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-6

```

RESULT 5
US-08-482-282A-4
; Sequence 4, Application US/08482282A
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy B.
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE
; TITLE OF INVENTION: CYSTEINE PROTEASE PROTEINS, NUCLEIC ACID

RESULT 6
US-08-768-619-2
; Sequence 2, Application US/08768619
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Griewe, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768,619
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,036
; FILING DATE: 07-JUN-1995

Query Match	63.4%;	Score 45;	DB 11;	Length 314;
Best Local Similarity	53.8%;	Pred. No. 6.4;		

```
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 1 KNSONKIIDLSP 13
Db 140 KTKNKLDSLSP 152

RESULT 8
US-08-768-619-22
; Sequence 22, Application US/08768619
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768,619
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,036
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1172
; US-08-768-619-25

Query Match 63.4%; Score 45; DB 11; Length 213;
Best Local Similarity 53.8%; Pred. No. 4.1;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSONKIIDLSP 13
Db 39 KTKNKLDSLSP 51

RESULT 9
US-08-768-619-25
; Sequence 25, Application US/08768619
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768,619
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,036
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1172
; US-08-768-619-25

Query Match 63.4%; Score 45; DB 11; Length 213;
Best Local Similarity 53.8%; Pred. No. 4.1;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSONKIIDLSP 13
Db 182 KTKNKLDSLSP 194

RESULT 10
US-08-768-619-33
; Sequence 33, Application US/08768619
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768,619
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,036
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1172
; US-08-768-619-25

Query Match 63.4%; Score 45; DB 11; Length 356;
Best Local Similarity 53.8%; Pred. No. 7.4;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSONKIIDLSP 13
Db 182 KTKNKLDSLSP 194

RESULT 10
US-08-768-619-33
; Sequence 33, Application US/08768619
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768,619
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,036
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1172
; US-08-768-619-25
```

; APPLICATION NUMBER: US/08/768,619
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,036
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-768-619-33

Query Match 63.4%; Score 45; DB 11; Length 400;
Best Local Similarity 53.8%; Pred. No. 8.5;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPE 13
| : : : : :
Db 224 KKTNKKLLDLSQP 236

RESULT 11

US-08-768-619-38
; Sequence 38, Application US/08768619
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/768,619
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,036
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: protein
; US-08-768-619-38

Query Match 63.4%; Score 45; DB 11; Length 215;
Best Local Similarity 53.8%; Pred. No. 4.1;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPE 13
| : : : : :
Db 39 KKTNKKLLDLSQP 51

RESULT 12

US-08-768-619-44
; Sequence 44, Application US/08768619
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
; TITLE OF INVENTION: PROTEASE PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/768,619
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/486,036
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-33-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-768-619-44

Query Match 63.4%; Score 45; DB 11; Length 401;
Best Local Similarity 53.8%; Pred. No. 8.5;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPE 13
| : : : : :
Db 225 KKTNKKLLDLSQP 237

RESULT 13
US-08-768-619-46
; Sequence 46, Application US/08768619
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.

APPLICANT: Wisniewski, Nancy
APPLICANT: Grieve, Robert B.
APPLICANT: Frank, Glenn R.
TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/768,619
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/486,036
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-33-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-768-619-46

Query Match 63.4%; Score 45; DB 11; Length 241;
Best Local Similarity 53.8%; Pred. No. 4.7;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSPE 13
| : : : : :
Db 65 KTKNKKLLDSPQ 77

RESULT 14
US-09-005-298-2
Sequence 2, Application US/09005298
GENERAL INFORMATION:
APPLICANT: Tripp, Cynthia A.
APPLICANT: Wisniewski, Nancy
APPLICANT: Grieve, Robert B.
APPLICANT: Frank, Glenn R.
TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,298
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/768,619
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-33-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-005-298-2

Query Match 63.4%; Score 45; DB 14; Length 398;
Best Local Similarity 53.8%; Pred. No. 8.4;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSPE 13
| : : : : :
Db 224 KTKNKKLLDSPQ 236

RESULT 15
US-09-005-298-4
Sequence 4, Application US/09005298
GENERAL INFORMATION:
APPLICANT: Tripp, Cynthia A.
APPLICANT: Wisniewski, Nancy
APPLICANT: Grieve, Robert B.
APPLICANT: Frank, Glenn R.
TITLE OF INVENTION: NOVEL FILARIID NEMATODE CYSTEINE
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,298
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/768,619
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-33-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 314 amino acids
TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-005-298-4

Query Match 63.4%; Score 45; DB 14; Length 314;
Best Local Similarity 53.8%; Pred. NO. 6.4;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPE 13
| : : : : :
Db 140 KTKNKLLDLSFQ 152

Search completed: November 13, 1999, 05:08:35
Job time: 10861 sec

C>Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 10-Oct-1997
C:Accession: D64201
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
C.A.; Venter, J.C
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346
A:Accession: D64201
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-273 <TIGR>
A:Cross-references: GB:U39680; GB:I43967; NID:g1045681; PID:g1045682; TIGR:MG013
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: methyltetrahydrofolate dehydrogenase (NAD+) homology
C:Keywords: folate biosynthesis; oxidoreductase
F:2-263/Domain: methyltetrahydrofolate dehydrogenase (NAD+) homology <MTFD>

Query Match 58.5%; Score 41.5; DB 2; Length 273;
Best Local Similarity 53.3%; Pred. No. 5.9;
Matches 8; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

Qy 1 KNSQNKII-DLSPE 14
: ||||| :|||:
Db 217 RNKQKLCGDINPEG 231
:

RESULT 4
C70217
outer surface protein homolog - Lyme disease spirochete plasmid B/cp26
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
C:Accession: C70217
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kurlav, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: C70217
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-365 <KLE>
A:Cross-references: GB:AE000792; NID:g2689884; PID:g2689897; TIGR:BBB07
A:Experimental source: strain B51
C:Genetics:
A:Genome: plasmid

Query Match 57.7%; Score 41; DB 2; Length 365;
Best Local Similarity 61.5%; Pred. No. 9.9;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KNSQNKII-DLSPE 13
: ||||| :|||:
Db 57 KNGMKPIIDVSP 69
:

RESULT 5
S73509
tRNA (guanine-N1)-methyltransferase trmD - Mycoplasma pneumoniae (ATCC 29342) (SGC3)
N:Alternate names: hypothetical protein K05_orf210
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C>Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
C:Accession: S73509
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirk, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S73327; MUID:97105885
A:Accession: S73509
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-210 <HIM>
A:Cross-references: EMBL:AE000019; GB:U00089; NID:g1673839; PID:g1673846
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: trmD
A:Genetic code: SGC3
C:Superfamily: tRNA (guanine-N1) methyltransferase

Query Match 56.3%; Score 40; DB 2; Length 210;
Best Local Similarity 50.0%; Pred. No. 8.2;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KNSQNKII-DLSPE 14
: ||||| :|||:
Db 54 RTPQSKVLLSPE 67
:

RESULT 6
S47748
53.1K protein (kdgk-dctA intergenic region) precursor - Escherichia coli
N:Alternate names: hypothetical protein f498
C:Species: Escherichia coli
C>Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 14-Nov-1997
C:Accession: S47748; B65151
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47748
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-498 <PLU>
A:Cross-references: EMBL:U00039; NID:g466582; PID:g466665
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: B65151
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-498 <BLAT>
A:Cross-references: GB:AE000429; GB:U00096; NID:g2367235; PID:g1789946; UWGP:b3527
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yhjJ

Query Match 56.3%; Score 40; DB 2; Length 498;
Best Local Similarity 38.5%; Pred. No. 21;
Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KNSQNKII-DLSPE 13
: ||||| :|||:
Db 403 RSLQNVVDIAPE 415
:

RESULT 7
H69157
exonuclease ABC chain A - Methanobacterium thermoautotrophicum (strain Delta H)
N:Alternate names: exonuclease ABC (EC 3.1.1.-), chain A; uvrA protein
C:Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 19-Feb-1999
C:Accession: H69157
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.
; Olu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu

A:Reference number: A69000; MUID:98037514
A:Accession: H69157
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-962 <MTH>
A:Cross-references: GB:AE000828; GB:AE000666; NID:g2621504; PID:g2621509
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH443
A:Start codon: TTG
C:Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C:Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; P-loop
F:38-45/Region: nucleotide-binding motif A (P-loop)
F:632-915/Domain: ATP-binding cassette homology <ABCE>
F:649-656/Region: nucleotide-binding motif A (P-loop)

Query Match 54.9%; Score 39; DB 2; Length 962;
Best Local Similarity 66.7%; Pred. No. 64;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 SQNKIIDLSPEG 14
I : ||||| |||
Db 905 SADHIIDLSPEG 916

RESULT 8
G71475
Probable phosphoesterase (EC 3.1.-.-) - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998
C:Accession: G71475
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, S.; Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
A:Reference number: A71570; MUID:99000809
A:Accession: G71475
A:Molecule type: DNA
A:Residues: 1-290 <ARN>
A:Cross-references: GB:AE001347; GB:AE001273; NID:g3329210; PID:g3329215
A:Experimental source: serotype D, strain UW-3/Cx
C:Comment: This sequence has motifs characteristic of a variety of phosphoesterases.
C:Genetics:
A:Gene: icc
C:Superfamily: unassigned probable phosphoesterases; phosphoesterase core homology
C:Keywords: hydrolase
F:10-107/Domain: phosphoesterase core homology <PEC>

Query Match 54.9%; Score 39; DB 2; Length 290;
Best Local Similarity 61.5%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 NSQNKIIDLSPEG 14
I : : ||||| |||
Db 248 NARFHIIDLSPEG 260

RESULT 9
S74748
Hypothetical protein slr0981 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
C:Accession: S74748
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, N. K.; DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.

A:Reference number: S74322; MUID:97061201
A:Accession: S74748
A:Status: preliminary

Best Local Similarity 50.0%; Pred. No. 78;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KNSONKIIDLSP 12
:|||||
Db 85 RNLQNNVVDLRP 96

RESULT 12
F64349
hypothetical protein MJ0398 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997
C:Accession: F64349
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
: Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
:son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999
A:Accession: F64349
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-271 <BUL>
A:Cross-references: GB:U67492; GB:L77117; NID:gl591096; PID:gl591103; TIGR:MJ0398; PID:g
C:Genetics:
A:Map position: FOR359111-359926

Query Match 54.2%; Score 38.5; DB 2; Length 271;
Best Local Similarity 58.8%; Pred. No. 20;
Matches 10; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

QY 1 KNSON----KIIDLSP 12
||| | | | | |
Db 54 KNSNPNFKTKFIDLSP 70

RESULT 13
A45456
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain NQ03 - Paracoccus denitrificans
C:Species: Paracoccus denitrificans
C>Date: 24-Feb-1994 #sequence_revision 12-Apr-1996 #text_change 04-Sep-1998
C:Accession: S23948; A45456
R:Xu, X.; Matsuno-Yagi, A.; Yagi, T.
Arch. Biochem. Biophys. 296, 40-48, 1992
A:Title: Structural features of the 66-kDa subunit of the energy-transducing NADH-ubiqui
A:Reference number: S23946; MUID:92296779
A:Accession: S23948
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-673 <XUA>
A:Cross-references: GB:M84572; NID:gl50599; PID:gl50601
R:Xu, X.; Matsuno-Yagi, A.; Yagi, T.
Biochemistry 32, 968-981, 1993
A:Title: DNA sequencing of the seven remaining structural genes of the gene cluster enc
A:Reference number: A45456; MUID:93136200
A:Accession: A45456
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 658-673 <XUL>
A:Note: sequence extracted from NCBI backbone (NCBIN:123409, NCBI:123410)
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 11
C:Keywords: NAD; oxidoreductase

Query Match 53.5%; Score 38; DB 2; Length 673;
Best Local Similarity 61.5%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 NSQNKIIDLSP 14
| | | | | | | |

Db 199 NMQGNIIDLCPVG 211

RESULT 14
S04781
excinuclease ABC chain A - Micrococcus luteus
N:Alternate names: excision endonuclease ABC (EC 3.1.-.-), chain A; uvrA protein
C:Species: Micrococcus luteus, Micrococcus lysodeikticus
A:Variety: ATCC 4698
C>Date: 28-Feb-1990 #sequence_revision 30-Sep-1991 #text_change 17-Jul-1998
C:Accession: S04781
R:Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A:Title: Micrococcus luteus homolog of the Escherichia coli uvrA gene: identification
A:Reference number: S04781; MUID:89364717
A:Accession: S04781
A:Molecule type: DNA
A:Residues: 1-992 <SHI>
A:Cross-references: EMBL:X15867; NID:g44452; PID:g581347
A:Note: the authors translated the codon AAC for residue 467 as Asp, CAG for residue
C:Genetics:
A:Gene: uvrA
A:Start codon: GTG
C:Function:
A:Description: has ATPase and DNA binding activity; involved in DNA repair
C:Superfamily: excinuclease ABC chain A; ATP-binding cassette homology
C:Keywords: ATP; DNA binding; DNA repair; duplication; hydrolase; P-loop
F:62-69/Region: nucleotide-binding motif A (P-loop)
F:675-958/Domain: ATP-binding cassette homology <ABCE>
F:692-699/Region: nucleotide-binding motif A (P-loop)

Query Match 53.5%; Score 38; DB 2; Length 992;
Best Local Similarity 58.3%; Pred. No. 99;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 SONKIIDLSP 14
| : | | | | | |
Db 948 SADHVIDLGPEG 959

RESULT 15
S25820
dynamitin-related protein VPS1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: interferon-regulated resistance protein homolog; probable GTP-bind
C:Species: Saccharomyces cerevisiae
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Jan-1999
C:Accession: S25820; S38070; S13053; S14224; S29366; A35627
R:Duesterhoeft, A.; Philippsen, P.
Yeast 8, 749-759, 1992
A:Title: DNA sequencing and analysis of a 24.7 kb segment encompassing centromere CEN
A:Reference number: S25322
A:Accession: S25820
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-704 <DUE>
A:Cross-references: EMBL:X65124; NID:g3517; PID:g3528
A:Experimental source: strain S288C
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1992
R:Duesterhoeft, A.; Moestl, D.; Pohlmann, R.; Philippsen, P.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37811
A:Accession: S38070
A:Molecule type: DNA
A:Residues: 1-704 <DU2>
A:Cross-references: EMBL:Z28226; NID:g486404; PID:g486405; MIPS:YKR001c
R:Rothman, J.H.; Raymond, C.K.; Gilbert, T.; O'Hara, P.J.; Stevens, T.H.
Cell 61, 1063-1074, 1990
A:Title: A putative GTP binding protein homologous to interferon-inducible Mx protein
A:Reference number: A35627; MUID:90275602
A:Accession: S13053
A:Molecule type: DNA

A:Residues: 1-32, 'N', '34-110, 'E', '112-704 <ROT>
A:Cross-references: EMBL:M33315; NID:g173182; PID:g173183
A:Note: the authors translated the codon GAA for residue 111 as Asn
R:Bloom, K.S.
submitted to the EMBL Data Library, August 1990
A:Reference number: S14224
A:Accession: S14224
A:Molecule type: DNA
A:Residues: 1-140, 'O', '142-704 <BLO>
A:Cross-references: EMBL:X54316; NID:g4527; PID:g4528
R:Yeh, E.; Driscoll, R.; Coltrera, M.; Olin, A.; Bloom, K.
Nature 349, 713-715, 1991
A:title: A dynamin-like protein encoded by the yeast sporulation gene SP015.
A:Reference number: S29366; MUID:91141584
A:Accession: S29366
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-140, 'O', '142-704 <YEH>
A:Cross-references: EMBL:X54316; NID:g4527; PID:g4528
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1990
C:Genetics:
A:Gene: SGD:VPS1; SP015
A:Cross-references: SGD:S0001709; MIPS:YKR001c
A:Map position: 11R
C:Superfamily: dynamin-related protein VPS1; dynamin/Vps1p homology
C:Keywords: GTP binding; P-loop; purine nucleotide binding
F:36-43/Region: nucleotide-binding motif A (P-loop)
F:394-473/Domain: dynamin/Vps1p homology <DVH>
F:42/Binding site: GTP (Lys) #status predicted

Query Match 53.5%; Score 38; DB 2; Length 704;
Best Local Similarity 53.8%; Pred. No. 68;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
OY 1 KNSQNKIIDLSPE 13
.Db 347 KKYQNELINLGPE 359

Search completed: November 13, 1999, 12:07:58
Job time: 2070 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:33:33 ; Search time 51.07 Seconds
(without alignments)
7.749 Million cell updates/sec

Title: US-08-913-430-6

Perfect score: 71

Sequence: 1 KNSQNKIDLSPEG 14

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	71	100.0	416	1	P46192 mycoplasma
2	47	66.2	291	1	MLER_LACLA
3	41.5	58.5	273	1	FOLD MYCGE
4	40	56.3	210	1	TRMD MYCPN
5	40	56.3	498	1	YHJ1_ECOLI
6	39	54.9	278	1	ATND_MOUSE
7	39	54.9	471	1	RNT1_YEAST
8	39	54.9	1125	1	YE62_SCHPO
9	38.5	54.2	271	1	Y398_METJA
10	38	53.5	421	1	AIAT_PIG
11	38	53.5	672	1	NQ03_PARDE
12	38	53.5	992	1	UVRA_MICLU
13	38	53.5	704	1	VPS1_YEAST
14	38	53.5	3655	1	YAMB_SCHPO
15	37	52.1	280	1	ATNC_CHICK
16	37	52.1	279	1	ATND_CAVPO
17	37	52.1	279	1	ATND_HUMAN
18	37	52.1	277	1	ATND_XENLA
19	37	52.1	233	1	MIP_LEGPN
20	36	50.7	346	1	ANX1_BOVIN
21	36	50.7	345	1	ANX1_HUMAN
22	36	50.7	341	1	ANX1_PIG
23	36	50.7	571	1	CRT1_MAIZE
24	36	50.7	1328	1	EXO2_SCHPO
25	36	50.7	452	1	GAD_HUMAN
26	36	50.7	449	1	GAD_MOUSE
27	36	50.7	449	1	GAD_RAT
28	36	50.7	318	1	IFR_CICAR
29	36	50.7	520	1	NIFK_KLEPN
30	36	50.7	1245	1	POLS_SINDV
31	36	50.7	1245	1	POLS_SINDV
32	36	50.7	659	1	RECQ_HUMAN
33	36	50.7	1992	1	TR12_HUMAN
34	36	50.7	926	1	UVRA_AQUAE
35	36	50.7	1015	1	UVRA_DEIRA
36	36	50.7	952	1	UVRA_MYCCE
37	36	50.7	948	1	UVRA_MYCPN
38	36	50.7	972	1	UVRA_MYCTU
39	36	50.7	970	1	UVRA_SYNY3
40	36	50.7	569	1	UVRA_VITST
41	36	50.7	515	1	VAC1_YEAST
42	36	50.7	314	1	Y0BW_MYCTU
43	36	50.7	274	1	Y059_CAEEL

ALIGNMENTS

RESULT 1

P46_MYCHY STANDARD; PRT; 416 AA.
AC P46192;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 46 KD SURFACE ANTIGEN PRECURSOR (P46).
OS MYCOPLASMA HYOPNEUMONIAE
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 67-74 AND 338-358.
RC STRAIN-ATCC 25934 / J;
RX MEDLINE; 95204368.
RA FUTO S., SETO Y., MITSUSE S., MORI Y., SUZUKI T., KAWAI K.;
RT "Molecular cloning of a 46-kilodalton surface antigen (P46) gene from
Mycoplasma hyopneumoniae: direct evidence of CGG codon usage for
arginine.";
RL J. BACTERIOL. 177:1915-1917(1995).
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID ANCHOR
(POTENTIAL).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC
CC EMBL; D16682; G1220114; ALT INIT.
DR PROSITE; P500013; PROKAR_LIPOPROTEIN; 1.
DR PFAM; PF00532; Peripla_BP-like; 1.
KW ANTIGEN; MEMBRANE; LIPOPROTEIN; SIGNAL.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 416 46 KD SURFACE ANTIGEN.
FT LIPID 28 28 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 416 AA; 45401 MW; 3EA04FED CRC32;

Query Match 100.0%; Score 71; DB 1; Length 416;
Best Local Similarity 100.0%; Pred. No. 2.8e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNSQNKIDLSPEG 14

|||||
DB 229 KNSQNKIDLSPEG 242

RESULT 2

MLER_LACLA STANDARD; PRT; 291 AA.
ID MLER_LACLA
AC P16400;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DE 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE MALOLACTIC FERMENTATION SYSTEM TRANSCRIPTIONAL ACTIVATOR.
GN MLER.
OS LACTOCOCCUS LACTIS (SUBSP. LACTIS) (STREPTOCOCCUS LACTIS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC LACTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89255069.
RA RENAULT P., GAILLARDIN C., HESLOT H.;

Db 54 RTPQSKVLLSPEG 67

RESULT 5

YHJJ_ECOLI
ID YHJJ_ECOLI STANDARD; PRT; 498 AA.
AC P37648;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 53.1 KD PROTEIN IN KDGR-DCTA INTERGENIC REGION PRECURSOR.
GN YHJJ
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
CC [1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 94316500.
RA SOPHIA H.J., BURLAND V., DANIELS D.L., PLUNKETT G. III, BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL NUCLEIC ACIDS RES. 22:2576-2586(1994).
RN [2]
RP SEQUENCE OF 25-36.
RC STRAIN-K12 / EMG2;
RX MEDLINE; 97443975.
RA LINK A.J., ROBISON K., CHURCH G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12."
RL ELECTROPHORESIS 18:1259-1313(1997).
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
CC -1- HAS LOST THE ACTIVE SITE RESIDUES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16; ALSO KNOWN AS THE
CC INSULINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U00039; G466665; -
DR EMBL; AE000429; G1789946; -
DR ECOGENE; EGI2254; YHJJ
DR PROSITE; PS00143; INSULINASE; FALSE_NEG.
DR PFAM; PF00675; Insulinase; 1.
KW HYDROLASE; METALLOPROTEASE; ZINC; PERIPLASMIC; SIGNAL.
FT SIGNAL 1 24
FT CHAIN 25 498 PROTEIN YHJJ.
FT SEQUENCE 498 AA; 55527 MW; 16842D4E CRC32;
SQ

Query Match 56.3%; Score 40; DB 1; Length 498;

Best Local Similarity 38.5%; Pred. No. 12;

Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSP 13

DB 403 RSLQNVVDIAPE 415

RESULT 6

ATND_MOUSE
ID ATND_MOUSE STANDARD; PRT; 278 AA.
AC P97370;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SODIUM/POTASSIUM-TRANSPORTING ATPASE BETA-3 CHAIN (EC 3.6.1.37)
DE (SODIUM/POTASSIUM-DEPENDENT ATPASE) (ATPB-3).

GN ATP1B3.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-RETINA;
RX MEDLINE; 97157071.
RA BESIRLI C.G., GONG T.-W.L., LOMAX M.I.;
RT "Novel beta 3 isoform of the Na,K-ATPase beta subunit from mouse
RT retina."
RL BIOCHIM. BIOPHYS. ACTA 1350:21-26(1997).
CC -1- FUNCTION: THIS IS THE NON-CATALYTIC COMPONENT OF THE ACTIVE
CC ENZYME, WHICH CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE
CC EXCHANGE OF NA AND K IONS ACROSS THE PLASMA MEMBRANE. THE EXACT
CC FUNCTION OF THIS GLYCOPROTEIN IS NOT KNOWN.
CC -1- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA
CC AND GAMMA.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -1- THE BETA SUBUNIT SEEMS TO BE ENCODED BY A MULTIGENE FAMILY. EACH
CC DIFFERENT SUBUNIT MAY HAVE SPECIALIZED FUNCTIONS.
CC -1- SIMILARITY: BELONGS THE THE NA+/K+ AND H+ ATPASES BETA CHAIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U59761; G1762432; -
DR MGD; MGI-107788; ATP1B3.
DR PROSITE; PS00390; ATPASE_NA_K_BETA_1; 1.
DR PFAM; PF00287; Na,K-ATPase; 1.
KW SODIUM/POTASSIUM TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; SIGNAL-ANCHOR.
FT DOMAIN 1 38 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 39 66 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 67 278 EXTRACELLULAR (POTENTIAL).
FT DISULFID 128 144 BY SIMILARITY.
FT DISULFID 154 170 BY SIMILARITY.
FT CARBOHYD 191 249 BY SIMILARITY.
FT CARBOHYD 124 124 POTENTIAL.
FT CARBOHYD 197 197 POTENTIAL.
SQ SEQUENCE 278 AA; 31775 MW; 0412E9C5 CRC32;

Query Match 54.9%; Score 39; DB 1; Length 278;

Best Local Similarity 70.0%; Pred. No. 9.3;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 NKIIDLSPEG 14

DB 176 NRIDLPDG 185

RESULT 7

RNT1_YEAST
ID RNT1_YEAST STANDARD; PRT; 471 AA.
AC Q02555; Q04008; Q05038;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBONUCLEASE III (EC 3.1.26.3) (RNAse III).
GN RNT1 OR YMR239C OR YM9408.01C OR YM9959.21.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE; 96180724.
RA ELELA S.A., IGEL H., ARES M. JR.;
RT "Rnase III cleaves eukaryotic preribosomal RNA at a U3
RT snORNP-dependent site.";
RL CELL 85:115-124(1996).
RN [2]
RP SEQUENCE OF 1-111 FROM N.A.
RC STRAIN-S288C / AB972;
RA GENTLES S., BOWMAN S., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 110-471 FROM N.A.
RC STRAIN-S288C / AB972;
RA SKELTON J., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: DSRNA-SPECIFIC NUCLEASE THAT CLEAVES EUKARYOTIC
CC PRERIBOSOMAL RNA AT THE U3 SNORNP-DEPENDENT A0 SITE IN THE 5'
CC EXTERNAL TRANSCRIBED SPACER (ETS) AND IN THE 3' ETS. IN VITRO,
CC CLEAVES SYNTHETIC 5' ETS RNA A0 SITE IN THE ABSENCE OF SNORNA OR
CC OTHER FACTORS. HAS AN ESSENTIAL GROWTH FUNCTION IN ADDITION TO
CC PRE-RRNA PROCESSING.
CC -1- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE TO 5'-PHOSPHO-
CC MONESTER.
CC -1- SIMILARITY: BELONGS TO THE RNASE III FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U27016; G1185416; -
CC EMBL; Z48756; G736305; -
CC EMBL; Z49939; G887620; -
CC SGD; L0002844; RNT1.
CC PROSITE; PS00517; RIBONUCLEASE_III; 1.
CC PFAM; PF00035; dsrm; 1.
CC PFAM; PF00636; Ribonuclease 3; 1.
CC HYDROLASE; NUCLEASE; ENDONUCLEASE.
SQ SEQUENCE 471 AA; 54070 MW; 3FB4B26D CRC32;

Query Match 54.9%; Score 39; DB 1; Length 471;
Best Local Similarity 80.0%; Pred. No. 17;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 SONKIIDLSP 12
DB 62 SYNKIIDLSP 71

RESULT 8
ID YE62_SCHPO STANDARD; PRT; 1125 AA.
AC O14248;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 127.8 KD SERINE-RICH PROTEIN C6G10.02C IN CHROMOSOME I.
GN SPAC6G10.02C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
CC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: SOME, TO YEAST YGR238C.
CC

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z98603; E334322; -
CC PFAM; PF01344; Kelch; 3.
CC HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1125 AA; 127754 MW; 5C966FCF CRC32;

Query Match 54.9%; Score 39; DB 1; Length 1125;
Best Local Similarity 58.3%; Pred. No. 43;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 NSQNKIIDLSP 13
DB 1049 NSQOKFLSLSPQ 1060

RESULT 9
ID Y398_METJA STANDARD; PRT; 271 AA.
AC Q57841;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN MJ0398.
GN MJ0398.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
CC METHANOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLOUEK A.,
RA SCOTT J.L., GEORGEAN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL SCIENCE 273:1058-1073(1996).
CC -1- SIMILARITY: SOME, TO M.JANNASCHII MJ0314.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67492; G1591103; -
CC TIGR; MJ0398; -
CC HYPOTHETICAL PROTEIN.
SQ SEQUENCE 271 AA; 32120 MW; F07C4781 CRC32;

Query Match 54.2%; Score 38.5; DB 1; Length 271;
Best Local Similarity 58.8%; Pred. No. 11;
Matches 10; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

QY 1 KNSQN-----KIIDLSP 12
DB 54 KNSNPNFKTKFIDLSP 70

```

RESULT 10
ID ALAT_PIG STANDARD; PRT; 421 AA.
AC P50417;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ALPHA-1-ANTITRYPSIN PRECURSOR (ALPHA-1 PROTEASE INHIBITOR) (ALPHA-1-
DE ANTIPTROTEINASE).
GN PI.
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 97009792.
RA ARCHIBALD A.L., COOPERWHITE S., MELLINK C.H.M., LAHBIB-MANSAIS Y.,
RA GELLIN J.;
RT "Porcine alpha-1-antitrypsin (PI): cDNA sequence, polymorphism and
RT assignment to chromosome 7q2.4-q2.6.";
RL ANIM GENET 27:85-89(1996)
CC -!- FUNCTION: INHIBITOR OF SERINE PROTEASES. ITS PRIMARY TARGET IS
CC ELASTASE, BUT IT ALSO HAS A MODERATE AFFINITY FOR PLASMIN AND
CC THROMBIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; X88780; G975230; .
DR PROSITE; PS00284; SERPIN; 1.
DR PFAM; PF00079; serpin; 1.
DR HSSP; P01009; 8AP1.
KW SERPIN; SERINE PROTEASE INHIBITOR; GLYCOPROTEIN; PLASMA; SIGNAL.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 421 ALPHA-1-ANTITRYPSIN.
FT ACT_SITE 385 386 REACTIVE BOND.
FT CARBOHYD 73 73 POTENTIAL.
FT CARBOHYD 110 110 POTENTIAL.
FT SEQUENCE 421 AA; 47194 MW; B6A0758A CRC32;
SQ
Query Match 53.5%; Score 38; DB 1; Length 421;
Best Local Similarity 61.5%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 KNSQKIIDLSP 13
Db 190 KSGQKIVDLVDE 202
RESULT 11
ID NQ03_PARDE STANDARD; PRT; 672 AA.
AC P29915;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 3 (EC 1.6.5.3) (NADH
DE DEHYDROGENASE 1, CHAIN 3) (NDH-1, CHAIN 3).
GN NQ03.
OS PARACOCUS DENITRIFICANS.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;
OC PARACOCUS.

RESULT 12
ID UVRA_MICLU STANDARD; PRT; 992 AA.
AC P13567;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE EXCINUCLEASE ABC SUBUNIT A.
GN UVRA.
OS MICROCOCCUS LUTEUS (MICROCOCCUS LYSOIDEIKTICUS).
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; MICROCOCCACEAE; MICROCOCCUS.
OC ACTINOMYCETALES; MICROCOCCINEAE; MICROCOCCACEAE; MICROCOCCUS.

[1]
SEQUENCE FROM N.A.
MEDLINE: 89364717.
SHIOTA S., NAKAYAMA H.;
Identification of a mutation in the UV-sensitive mutant DB7.;
MOL. GEN. GENET. 217:332-340(1989).
-1- FUNCTION: THIS IS ONE OF THE THREE SUBUNITS OF THE ABC EXCISION
NUCLEASE, A DNA REPAIR ENZYME THAT CATALYZES THE EXCISION
REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS PRODUCING OLIGOMERS
HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE AND A DNA-BINDING
PROTEIN THAT PREFERENTIALLY BINDS SINGLE-STRANDED OR UV-IRRADIATED
DOUBLE-STRANDED DNA.
-1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
(ABC TRANSPORTERS). CONTAINS TWO ABC DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X15867; G581347; -
PIR: S04781; S04781.
PROSITE: PS00211; ABC_TRANSPORTER; 2.
PFAM: PF00005; ABC_tran; 2.
SOS RESPONSE; EXCISION NUCLEASE; DNA REPAIR; ATP-BINDING; REPEAT;
DNA-BINDING; ZINC-FINGER.
NP_BIND 62 69 ATP.
FT NP_BIND 692 699 ATP.
FT ZN_FING 302 330 C4-TYPE.
FT ZN_FING 791 817 C4-TYPE.
SEQUENCE 992 AA; 108692 MW; 833D981F CRC32;

Query Match 53.5%; Score 38; DB 1; Length 992;
Best Local Similarity 58.3%; Pred. No. 57;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 SONKIIDLSPG 14
| : : : : :
Db 948 SADHVIDLGPEG 959

RESULT 13
VPS1_YEAST
ID VPS1_YEAST STANDARD; PRT; 704 AA.
AC P21576;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE VACUOLAR SORTING PROTEIN 1.
GN VPS1 OR SPO15 OR LAM1 OR YKR001C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
[1]
SEQUENCE FROM N.A.
RC STRAIN-J17;
RX MEDLINE: 91141584.
RA YEH E.Y., DRISCOLL R., COLTRERA M., OLINS A., BLOOM K.S.;
RT "A dynamin-like protein encoded by the yeast sporulation gene SPO15.";
RL NATURE 349:713-715(1991).
[2]
SEQUENCE FROM N.A.
RC STRAIN-J17;
RX MEDLINE: 90275602.
RA ROTHMAN J.H., RAYMOND C.K., GILBERT T., O'HARA P.J., STEVENS T.H.;
RT "A putative GTP binding protein homologous to interferon-inducible Mx
proteins performs an essential function in yeast protein sorting.";
RL CELL 61:1063-1074(1990).

[3]
SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE: 93070612.
RA DUESTERHOEF A., PHILIPSEN P.;
RT "DNA sequencing and analysis of a 24.7 kb segment encompassing
centromere CEN11 of Saccharomyces cerevisiae reveals nine previously
unknown open reading frames.";
YEAST 8:749-759(1992).
[4]
CHARACTERIZATION.
RX MEDLINE: 93345450.
RA WILSBACH K., PAYNE G.S.;
RT "Vpslp, a member of the dynamin GTPase family, is necessary for Golgi
membrane protein retention in Saccharomyces cerevisiae.";
EMBO J. 12:3049-3059(1993).
-1- FUNCTION: ESSENTIAL FOR PROTEIN SORTING IN MEIOTIC CELL DIVISION
OF SACCHAROMYCES CEREVISIAE; IT BINDS MICROTUBULES. COULD ALSO BE
INVOLVED IN MICROTUBULE-ASSOCIATED MOTILITY. NECESSARY FOR
MEMBRANE PROTEIN RETENTION IN A LATE GOLGI COMPARTMENT. INTERACTS
WITH THE MVPL PROTEIN.
-1- SUBCELLULAR LOCATION: CYTOPLASMIC.
-1- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X54316; G4528; -
DR EMBL: X33315; G173183; -
DR EMBL: X65144; G3528; -
DR EMBL: Z28226; G486405; -
DR PIR: S25820; S25820.
DR SGD: L0002006; SPO15.
DR PROSITE: PS00410; DYNAMIN; 1.
DR PFAM: PF00350; dynamin; 2.
DR PFAM: PF01031; dynamin_2; 1.
KW MOTOR PROTEIN; MEIOSIS; CELL DIVISION; MICROTUBULES; GTP-BINDING.
FT NP_BIND 36 43
FT NP_BIND 178 182 GTP (POTENTIAL).
FT NP_BIND 247 250 GTP (POTENTIAL).
FT CONFLICT 33 33 T -> N (IN REF. 2).
FT CONFLICT 111 111 N -> E (IN REF. 2).
FT CONFLICT 141 141 K -> Q (IN REF. 1).
SEQUENCE 704 AA; 78736 MW; F676B06B CRC32;

Query Match 53.5%; Score 38; DB 1; Length 704;
Best Local Similarity 53.8%; Pred. No. 39;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KNSQKIIDLSPE 13
| : : : : :
Db 347 KXQNELINLGPE 359

RESULT 14
YAMB_SCHPO
ID YAMB_SCHPO STANDARD; PRT; 3655 AA.
AC Q10064;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 420.8 KD PROTEIN C1F5.11C IN CHROMOSOME 1.
GN SPAC1F5.11C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.

```

RN SEQUENCE FROM N.A.
RP STRAIN=972;
RA GENTLES S., ODELL C., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A.,
RA WALSH S.V.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: STRONG, TO YEAST YHR099W.
CC -1- SIMILARITY: BELONGS TO THE P13/P14-KINASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: Z68136; E212003;
KW HYPOTHETICAL PROTEIN; TRANSFERASE; KINASE.
FT DOMAIN 3324 3655 P13K/PI4K (BY SIMILARITY).
SQ SEQUENCE 3655 AA; 420774 MW; 6B082A74 CRC32;

Query Match 53.58; Score 38; DB 1; Length 3655;
Best Local Similarity 61.58; Pred. No. 2.4e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 NSONKIIDLSPG 14
   | | | | | | |
Db 3070 SSDNKRDLSPG 3082

RESULT 15
ATNC CHICK
ID ATNC_CHICK STANDARD; PRT; 280 AA.
AC P33879;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SODIUM/POTASSIUM-TRANSPORTING ATPASE BETA-2 CHAIN (EC 3.6.1.37)
DE (SODIUM/POTASSIUM-DEPENDENT ATPASE).
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
[1]
SEQUENCE FROM N.A.
STRAIN-WHITE LEGHORN; TISSUE-BRAIN;
MEDLINE; 93312842.
RA LEMAS M., FAMBROUGH D.M.;
RT "Sequence analysis of DNA encoding an avian Na,K(+-)ATPase beta 2-
subunit."
RL BIOCHIM. BIOPHYS. ACTA 1149:339-342(1993).
CC -1- FUNCTION: THIS IS THE NON-CATALYTIC COMPONENT OF THE ACTIVE
CC ENZYME, WHICH CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE
CC EXCHANGE OF NA AND K IONS ACROSS THE PLASMA MEMBRANE. THE EXACT
CC FUNCTION OF THIS GLYCOPROTEIN IS NOT KNOWN.
CC -1- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA
CC AND GAMMA.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN.
CC -1- THE BETA SUBUNIT SEEMS TO BE ENCODED BY A MULTIGENE FAMILY. EACH
CC DIFFERENT SUBUNIT MAY HAVE SPECIALIZED FUNCTIONS.
CC -1- SIMILARITY: BELONGS THE THE NA+/K+ AND H+ ATPASES BETA CHAIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```

DR EMBL: L13208; G289819;
DR PROSITE; PS00390; ATPASE_NA_K_BETA_1; 1.
DR PROSITE; PS00391; ATPASE_NA_K_BETA_2; 1.
DR PFAM; PF00287; Na_K-ATPase; 1.
KW SODIUM/POTASSIUM TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; SIGNAL-ANCHOR.
FT DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 38 62 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 63 280 EXTRACELLULAR (POTENTIAL).
FT DISULFID 128 145 BY SIMILARITY.
FT DISULFID 153 171 BY SIMILARITY.
FT DISULFID 192 231 BY SIMILARITY.
FT CARBOHYD 96 96 POTENTIAL.
FT CARBOHYD 148 148 POTENTIAL.
FT CARBOHYD 241 241 POTENTIAL.
SQ SEQUENCE 280 AA; 31858 MW; 5E10C8D3 CRC32;

Query Match 52.18; Score 37; DB 1; Length 280;
Best Local Similarity 70.08; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 NKIIDLSPG 14
   | | | | |
Db 177 NRIIGLKPEG 186

Search completed: November 13, 1999, 10:33:35
Job time: 5185 sec

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:26 ; Search time 139.86 Seconds
(without alignments)
6.161 Million cell updates/sec

Title: us-08-913-430-6

Perfect score: 71

Sequence: 1 KNSQNKIIDLSPEG 14

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPTREMBL_10.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	63.4	400	5	O18348 dirofilaria
2	45	63.4	367	5	O16667 caenorhabdi
3	41	57.7	365	2	O50983 borrelia bu
4	40	56.3	227	2	O50188 mycoplasma
5	40	56.3	204	2	O30937 legionella
6	39	54.9	962	1	O26543 methanobact
7	39	54.9	290	2	O84759 chlamydia t
8	39	54.9	187	2	P72883 synchocyst
9	39	54.9	684	5	O16652 caenorhabdi
10	39	54.9	1159	5	O60981 leishmania
11	39	54.9	648	11	Q92129 mus musculus
12	39	54.9	631	11	Q92128 mus musculus
13	39	54.9	1245	12	Q9YJX7 archibis vir
14	38	53.5	46	1	O29540 caenaeoglob
15	38	53.5	204	2	O68491 legionella
16	38	53.5	232	2	O32763 legionella
17	38	53.5	233	2	O32839 legionella
18	38	53.5	234	2	O32759 legionella
19	38	53.5	233	2	O32753 legionella
20	38	53.5	401	4	O13302 homo sapien
21	38	53.5	419	4	Q14722 homo sapien
22	38	53.5	401	4	Q99869 homo sapien
23	38	53.5	408	4	Q16547 homo sapien
24	38	53.5	513	4	O75178 homo sapien
25	38	53.5	1484	5	O94298 caenorhabdi
26	38	53.5	408	6	O28528 mustela put
27	38	53.5	401	11	Q61763 mus musculus
28	38	53.5	401	11	O63277 rattus norv
29	38	53.5	401	11	P97380 mus musculus

30	37	52.1	204	2	O68489 legionella
31	37	52.1	204	2	O68492 legionella
32	37	52.1	828	2	O25068 helicobacte
33	37	52.1	1119	2	O51228 borrelia bu
34	37	52.1	731	2	O70039 streptococc
35	37	52.1	603	2	O70041 streptococc
36	37	52.1	210	2	O48552 legionella
37	37	52.1	210	2	O48881 legionella
38	37	52.1	202	2	O30934 legionella
39	37	52.1	199	2	O30935 legionella
40	37	52.1	204	2	O30936 legionella
41	37	52.1	204	2	O30938 legionella
42	37	52.1	204	2	O30939 legionella
43	37	52.1	202	2	O30940 legionella
44	37	52.1	204	2	O30941 legionella
45	37	52.1	204	2	O30942 legionella

ALIGNMENTS

```

RESULT 1
O18348
ID O18348 PRELIMINARY; PRT; 400 AA.
AC O18348;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE CSTEINE PROTEASE (FRAGMENT).
GN NDICP1298.
OS Dirofilaria immitis.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida;
OC Filarioidea; Onchocercidae; Dirofilaria.
RN [1]
RP SEQUENCE FROM N.A.
RA TRIPP C.A., GRIEVE R.B.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001101; AAB62536.1; -
DR PFAM; PF001112; Peptidase_C1; 1.
KW Protease.
FT NON_TER
SQ SEQUENCE 400 AA; 45768 MW; C8B8B81B CRC32;

```

```

Query Match 63.4%; Score 45; DB 5; Length 400;
Best Local Similarity 53.8%; Pred No. 4;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 KNSQNKIIDLSPE 13
| : : : : :
Db 224 KKTNKLIDLSPO 236

```

```

RESULT 2
O16667
ID O16667 PRELIMINARY; PRT; 367 AA.
AC O16667;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE T27B7.1 PROTEIN (FRAGMENT).
GN T27B7.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

```

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF016675; AAB66133.1; -;
 DR PFAM; PF01362; DUF12.1;
 DR PFAM; PF00105; zf-C4.1;
 FT NON_TER 367 367
 SQ SEQUENCE 367 AA; 41585 MW; 49A9BA0A CRC32;

Query Match 63.4%; Score 45; DB 5; Length 367;
 Best Local Similarity 90.0%; Pred. No. 3.7;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 SONKIIDLSP 12
 |||||
 DB 131 SONKFIDLSP 140

RESULT 3
 OS0983 PRELIMINARY; PRT; 365 AA.
 AC OS0983;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE OUTER SURFACE PROTEIN, PUTATIVE.
 GN BB807.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OG plasmid cp26.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35210 / B31;
 RX MEDLINE; 98065943.
 RA FRASER C.M., CASYENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
 RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKLEY E.K., GWINN M.,
 RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
 RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
 RA VAN VUUT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
 RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
 RA SMITH H.O., VENTER J.C.;
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
 burgdorferi.";
 RL Nature 390:580-586(1997).
 DR EMBL; AE000792; AAC66321.1; -;
 DR TIGR; BB807; -;
 KW Plasmid.
 SQ SEQUENCE 365 AA; 41916 MW; A5D3D43B CRC32;

Query Match 57.7%; Score 41; DB 2; Length 365;
 Best Local Similarity 61.5%; Pred. No. 18;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 KNSQNKIIDLSP 13
 |||||

Db 57 KNGMKPIIDVSPE 69
 |||||
 RESULT 4
 OS0188 PRELIMINARY; PRT; 227 AA.
 AC OS0188;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-JUN-1998 (Tremblrel. 06, Last annotation update)
 DE HYPOTHETICAL 26.3 KD PROTEIN.
 OS Mycoplasma hyopneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-232A;
 RA HSU T., MINION F.C.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF012905; AAC32528.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 227 AA; 26285 MW; 66FCFE1C CRC32;

Query Match 56.3%; Score 40; DB 2; Length 227;
 Best Local Similarity 61.5%; Pred. No. 17;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 NSQNKIIDLSP 14
 |||||
 DB 174 NPQNTIIIGPEG 186

RESULT 5
 OS0937 PRELIMINARY; PRT; 204 AA.
 AC OS0937;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE MACROPHAGE INFECTIVITY POTENTIATOR PROTEIN (FRAGMENT).
 GN MIP.
 OS Legionella pneumophila serogroup 5.
 OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae;
 OC Legionella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC33216;
 RA RATCLIFF R.M., LANSER J.A., MANNING P.A., HEUZENROEDER M.W.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF022319; AAB81367.1; -;
 DR PFAM; PF00254; FKBP; 1.
 DR PFAM; PF01346; FKBP.N; 1.
 DR PROSITE; PS00453; FKBP_PPIASE.1; 1.
 DR PROSITE; PS00454; FKBP_PPIASE.2; 1.
 FT NON_TER 1
 NON_TER 204 204
 SQ SEQUENCE 204 AA; 21655 MW; 5EC041F4 CRC32;

Query Match 56.3%; Score 40; DB 2; Length 204;
 Best Local Similarity 61.5%; Pred. No. 15;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KNSQNKIIDLSP 13
 |||||

RESULT 6
 OS26543 PRELIMINARY; PRT; 962 AA.
 ID OS26543

AC O25543; 1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE EXINUCLEASE ABC SUBUNIT A.
 GN MTH443
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RX MEDLINE; 98037514.
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
 RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
 RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,
 RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
 RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
 RA MCDONALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 J. Bacteriol. 179:7135-7155(1997)."
 DR EMBL; AE000828; AAB84949.1;
 DR PFAM; PF00005; ABC_tran; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
 KW ATP-binding; Transport.
 SQ SEQUENCE 962 AA; 108395 MW; ACEB62BC CRC32;
 Query Match 54.98; Score 39; DB 1; Length 962;
 Best Local Similarity 66.78; Pred. No. 1.2e+02;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 3 SONKIIDLSPEG 14
 I: | | | | | | | |
 DB 905 SADHIIDLGPEG 916
 RESULT 7
 O84759
 ID O84759 PRELIMINARY; PRT; 290 AA.
 AC O84759
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE (PHOSPHOHYDROLASE).
 GN ICC.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/UW-3/CX;
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
 RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
 RA DAVIS R.W.;
 RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
 Chlamydia trachomatis."
 RL Science 0:0-0(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/CX;
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
 RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
 RA DAVIS R.W.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE001347; AAC68349.1;
 KW Hydrolase.
 SQ SEQUENCE 290 AA; 32716 MW; 3248C5DF CRC32;
 Query Match 54.98; Score 39; DB 2; Length 290;
 Best Local Similarity 61.58; Pred. No. 32;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 NSQNKIIDLSPEG 14
 I: | | | | | | | |
 DB 248 NARPHIIDLHPG 260
 RESULT 8
 P72883
 ID P72883 PRELIMINARY; PRT; 187 AA.
 AC P72883
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
 DE HYPOTHETICAL 20.9 KD PROTEIN.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RA TABATA S.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC6803;
 RX MEDLINE; 97061201.
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
 RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M.,
 RA TABATA S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D90901; BAA16899.1;
 KW Hypothetical protein.
 SQ SEQUENCE 187 AA; 20887 MW; E74ED38D CRC32;
 Query Match 54.98; Score 39; DB 2; Length 187;
 Best Local Similarity 57.18; Pred. No. 20;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 KNSQNKIIDLSPEG 14
 I: | | | | | | | |
 DB 32 RNPQCKLIILSPVG 45
 RESULT 9
 O16652
 ID O16652 PRELIMINARY; PRT; 684 AA.
 AC O16652
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE T06D4.4 PROTEIN.
 GN T06D4.4
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

```

RA THIRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA LE T.T., GOELA D.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016673; AAB66123.1; -.
DR PFAM; PF01431; Peptidase_W13; 1.
SQ SEQUENCE 684 AA; 79176 MW; 248CEE1 CRC32;

Query Match 54.9%; Score 39; DB 5; Length 684;
Best Local Similarity 50.0%; Pred. No. 81;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSPE 14
Db 353 KHSHTLIDLLPRG 366

RESULT 10
ID O60981 PRELIMINARY; PRT; 1159 AA.
AC O60981;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
DE CAK2.
GN CAK2.
OS Leishmania major.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA MYLER P.J., AUDLEMAN L., HIXSON G., KISER P., LEMLEY C., RICKEL E.,
RA SISK E., SUNKIN S., SWARTZELL S., WESTLAKE T., MAGNESS C., BASTIEN P.,
RA FU G., IVENS A., STUART K.;
RT "Leishmania major Friedlin chromosome 1 has only two polycistronic
RT units of protein coding genes.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA MYLER P.J.;
RT "The nucleotide sequence of Leishmania major Friedlin chromosome 1.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE001274; AAC24689.1; -.
SQ SEQUENCE 1159 AA; 126733 MW; 30EE835 CRC32;

Query Match 54.9%; Score 39; DB 5; Length 1159;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSP 12
Db 85 RNLQNNVVDLRP 96

RESULT 11
ID O92129 PRELIMINARY; PRT; 648 AA.
AC O92129
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DR EMBL; AB017105; BAA75086.1; -.
SQ SEQUENCE 631 AA; 70608 MW; 78AB5F01 CRC32;

Query Match 54.9%; Score 39; DB 11; Length 631;
Best Local Similarity 61.5%; Pred. No. 74;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSP 13
Db 180 KNSQLKLIYVTP 192

RESULT 12
ID O92128 PRELIMINARY; PRT; 631 AA.
AC O92128;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE DNA HELICASE Q1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RA WANG W.S., SEKI M., YAMAOKA T., SEKI T., TADA S., KATADA T.,
RA FUJIMOTO H., ENOMOTO T.;
RT "Cloning of two isoforms of mouse DNA helicase Q1/RecQL cDNA; alpha
RT form is expressed ubiquitously and beta form specifically in the
RT testis.";
RL Blochim. Biophys. Acta 1443:198-202(1998).
DR EMBL; AB017104; BAA75085.1; -.
KW Helicase.
SQ SEQUENCE 648 AA; 72552 MW; 5598CED CRC32;

Query Match 54.9%; Score 39; DB 11; Length 648;
Best Local Similarity 61.5%; Pred. No. 76;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KNSQNKIIDLSP 13
Db 180 KNSQLKLIYVTP 192

RESULT 13
ID O92128 PRELIMINARY; PRT; 1245 AA.
AC O92128;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE STRUCTURAL POLYPROTEIN.
OS Sindbis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;

```

OC Alphavirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=XJ-160;
 RA Li L., Liang G.D., Zhou G.L., Fu S.H., Jin Q., Hou Y.D.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF103728; AAC83379.1; -
 KW Polypeptide.
 FT CHAIN 1 264 CAPSID PROTEIN.
 FT CHAIN 265 328 E3.
 FT CHAIN 329 751 E2.
 FT CHAIN 752 806 E1.
 FT CHAIN 807 1245 E1.
 SQ SEQUENCE 1245 AA; 136876 MW; DF961C94 CRC32;

Query Match 54.9%; Score 39; DB 12; Length 1245;
 Best Local Similarity 61.5%; Pred. No. 1.5e+02;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 NSQNKIIDLSPG 14
 ||: || ||||
 Db 248 NSKGKTIKTSPEG 260

RESULT 14
 O29540 PRELIMINARY; PRT; 46 AA.
 AC O29540;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
 DE HYPOTHETICAL 5.3 KD PROTEIN.
 GN AF0718.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE: 98049343.
 RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
 RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESSE C.R.,
 RA VENTER J.C.;
 RT "The complete genome sequence of the hyperthermophilic,
 RT sulphate-reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 DR EMBL; AE001055; AAB90531.1; -
 DR TIGR; AF0718; -
 KW Hypothetical protein.
 SQ SEQUENCE 46 AA; 5271 MW; FD00C65F CRC32;

Query Match 53.5%; Score 38; DB 1; Length 46;
 Best Local Similarity 60.0%; Pred. No. 6.8;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 NKIIDLSPG 14
 ||: || ||
 Db 10 NKVVDLRPPG 19

RESULT 15
 O68491 PRELIMINARY; PRT; 204 AA.
 ID O68491

AC O68491;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE MACROPHAGE INFECTIVITY POTENTIATOR PROTEIN (FRAGMENT).
 GN MIP.
 OS Legionella gormanii.
 OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae;
 OC Legionella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IMVS-LC0777C;
 RA RATCLIFF R.M., LANSEY J.A., MANNING P.A., HEUZENROEDER M.W.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF047748; AAC08972.1; -
 DR PFAM; PF00254; FKBP; 1.
 DR PFAM; PF01346; FKBP_N; 1.
 DR PROSITE; PS00453; FKBP_PPIASE_1; 1.
 DR PROSITE; PS00454; FKBP_PPIASE_2; 1.
 FT NON_TER 1
 FT NON_TER 204 204
 SQ SEQUENCE 204 AA; 21713 MW; 18FC0BA8 CRC32;

Query Match 53.5%; Score 38; DB 2; Length 204;
 Best Local Similarity 53.8%; Pred. No. 33;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNSQNKIIDLSPG 13
 ||: || ||:
 Db 38 KNFKNOGIDINPE 50

Search completed: November 13, 1999, 12:55:28
 Job time: 3027 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:07:58 ; Search time 75.45 seconds
(without alignments)
7.434 Million cell updates/sec

Title: US-08-913-430-7

Perfect score: 70

Sequence: 1 AGXWAKETTREKS 14

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	41	58.6	26926	1	I38344	titin, cardiac muscle
2	41	58.6	526	2	T02131	cytochrome P450 ho
3	40	57.1	765	1	ISHU71	DNA topoisomerase
4	39	55.7	485	2	C69584	aldehyde dehydroge
5	38	54.3	435	1	TPHUN1	protein-tyrosine-p
6	38	54.3	390	2	I38202	leupin precursor
7	37	53.6	511	2	C56849	dopamine receptor
8	37	52.9	1136	1	S78745	protein-tyrosine k
9	37	52.9	2330	1	RIWNV	genome polyprotein
10	37	52.9	767	2	A49546	DNA topoisomerase
11	37	52.9	390	2	I38201	scumous cell carc
12	37	52.9	458	2	C69421	methylviologen red
13	37	52.9	473	2	A70460	DNA polymerase III
14	36	51.4	935	1	I41193	outer membrane pro
15	36	51.4	362	2	S10512	phosphoserine tran
16	36	51.4	362	2	B64830	bone sialoprotein
17	36	51.4	310	2	I46987	myosin-light-chain
18	36	51.4	1004	2	A55142	DNA polymerase III
19	36	51.4	1443	2	S73446	hypothetical prote
20	36	51.4	197	2	A65081	hypothetical prote
21	36	51.4	195	2	D71872	probable membrane
22	36	51.4	141	2	S49955	Sec13 protein - ye
23	36	51.4	297	2	A45442	cell blood group p
24	35	50.0	732	1	HVHUK	dynein heavy chain
25	35	50.0	4644	1	A38905	suppressor 2 prote
26	35	50.0	685	1	EFBY52	phosphoserine tran
27	35	50.0	362	2	S71439	L-arabinose isomer
28	35	50.0	498	2	C69587	leucine--trNA liga
29	35	50.0	1090	2	JC1421	adenomatous polypo
30	35	50.0	2845	2	I49505	adenine deaminase
31	35	50.0	556	2	H69279	hvaA - Pseudomonas
32	35	50.0	690	2	B42594	surface-layer prot
33	35	50.0	917	2	S26365	hypothetical prote
34	35	50.0	770	2	F65028	hypothetical prote
35	35	50.0	224	2	A28988	hypothetical prote
36	35	50.0	1668	2	A60272	iga-specific metal
37	35	50.0	416	2	S16658	hypothetical prote
38	35	50.0	176	2	B70606	adenosylmethionine
39	35	50.0	361	2	T01934	

40	35	50.0	376	2	S45763	hypothetical prote
41	35	50.0	264	2	S49787	hypothetical prote
42	35	50.0	304	2	S65198	hypothetical prote
43	35	50.0	375	2	A39777	41k blood stage an
44	35	50.0	1333	2	S38635	blastopia polyprot
45	34.5	49.3	467	2	T02238	glucosyl transfera

ALIGNMENTS

RESULT 1

I38344

titin, cardiac muscle - human

N:Alternate names: Connectin

N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)

C:Species: Homo sapiens (man)

C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 02-Jul-1998

C:Accession: I38344; I38345; S20898; S20899; S63665; S37393

R:Labeit, S.; Kolmerer, B.

Science 270, 293-296, 1995

A:Title: titins: giant proteins in charge of muscle ultrastructure and elasticity.

A:Reference number: A57430; MUID:96026330

A:Accession: I38344

A>Status: nucleic acid sequence not shown; translation not shown; translated from GB/

A:Molecule type: mRNA

A:Residues: 1-26926 <LAB1>

A:Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425

R:Musco, G.; Tziatzios, C.; Schuck, F.; Pastore, A.

Biochemistry 34, 553-561, 1995

A:Title: Dissecting titin into its structural motifs: identification of an alpha-heli

A:Reference number: I38345; MUID:95119041

A:Accession: I38345

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1977-2014 <MUS>

A:Cross-references: EMBL:X83270; NID:g602579; PID:g602580

A>Note: conformation and properties are reported for a synthetic peptide correspondin

R:Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.

EMBO J. 11, 1711-1716, 1992

A:Title: Towards a molecular understanding of titin.

A:Reference number: S20897; MUID:92258380

A:Accession: S20898

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 13597-14200, 'I', 14202-14696 <LAB2>

A:Cross-references: EMBL:X64698; NID:g37192; PID:g37193

A:Accession: S20897

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>

A:Cross-references: EMBL:X64697; NID:g37190; PID:g37195

A:Accession: S20899

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 7, 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 2248

A:Cross-references: EMBL:X64697; NID:g37190; PID:g37195

R:Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.

J. Mol. Biol. 256, 556-563, 1996

A:Title: Genomic organization of M line titin and its tissue-specific expression in t

A:Reference number: S63665; MUID:96177761

A:Accession: S63665

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 26729-26825 <KOL>

A:Cross-references: EMBL:X92412; NID:g1236761

R:Gautel, M.; Leonard, K.; Labeit, S.

EMBO J. 12, 3827-3834, 1993

A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in different

A:Reference number: S37393; MUID:94008990

A:Accession: S37393

A:Molecule type: mRNA

A:Residues: 26831-26926 <GAU>

R:Improta, S.; Politou, A.S.; Pastore, A.
 submitted to the Brookhaven Protein Data Bank, February 1996
 A:Reference number: A66736; PDB:1TIT
 A:Contents: annotation: conformation by (1)H-NMR, residues 5253-5341
 R:Pfuhl, M.; Pastore, A.
 submitted to the Brookhaven Protein Data Bank, August 1996
 A:Reference number: A66201; PDB:1NCT
 A:Contents: annotation: conformation by (1)H-NMR, residues 'S', 26059-26155
 C:Genetics:
 A:Gene: GDB:TTN
 A:Cross-references: GDB:127867; OMIM:188840
 A:Map position: 2q31-2q32
 C:Function:
 A:Description: structural protein forming filaments in striated muscle
 C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro
 C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
 ral protein
 F:24752-25008/Domain: protein kinase homology <KIN>
 F:84.177.905.2276.2378.2459.2481.2563.2669.2763.2896.3088.3179.3384.3432.3628.3772.4068.
 98.11066.11488.11515.11635.11945.12170.12478.12526.12645.12875.13001.13036.13295.13540.1
 tus predicted
 F:16780.16976.17579.17602.17667.17681.17845.17899.18121.18188.18209.18336.18670.18680.18
 21900.21935.22295.22495.22627.22897.23024.23318.23883.24012.24177.24290.24447.24642.248
 F:26171.26178.26184.26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 58.6%; Score 41; DB 1; Length 26926;
 Best Local Similarity 70.0%; Pred. No. 1.1e+03;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETKEEK 13
 | | | | |
 Db 1995 WTKELTEEEK 2004

RESULT 2
 T02191
 cytochrome P450 homolog - Arabidopsis thaliana
 N:Alternate names: F14M4.21 protein
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Mar-1999
 C:Accession: T02191
 R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; Ron
 submitted to the EMBL Data Library, September 1998
 A:Description: Arabidopsis thaliana chromosome II BAC F14M4 genomic sequence.
 A:Reference number: 214609
 A:Accession: T02191
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-526 <ROU>
 A:Cross-references: EMBL:AC004411; NID:g3522932; PID:g3522945
 C:Genetics:
 A:Introns: 93/1; 168/3; 254/2; 375/3
 A:Note: F14M4.21

Query Match 58.6%; Score 41; DB 2; Length 526;
 Best Local Similarity 77.8%; Pred. No. 19;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 WAKETKEE 12
 | | | | |
 Db 185 WRKSTKEE 193

RESULT 3
 ISHUT1
 DNA topoisomerase (EC 5.99.1.2) - human
 N:Alternate names: nicking-closing enzyme; omega-protein; relaxing enzyme; swivelase; ty
 C:Species: Homo sapiens (man)
 C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Sep-1998
 C:Accession: A30887; A40008; S13821; S02397; S40643; A34422
 R:D'Arpa, P.; Machlin, P.S.; Rattie III, H.; Rothfield, N.F.; Cleveland, D.W.; Earnshaw,

Proc. Natl. Acad. Sci. U.S.A. 85, 2543-2547, 1988
 A:Title: cDNA cloning of human DNA topoisomerase I: catalytic activity of a 67.7-kDa
 A:Reference number: A30887; MUID:88190108
 A:Accession: A30887
 A:Molecule type: mRNA
 A:Residues: 1-765 <DAR>
 A:Cross-references: GB:J03250; NID:g339805; PID:g339806
 R:Kunze, N.; Yang, G.; Doelberg, M.; Sundarp, R.; Knippers, R.; Richter, A.
 J. Biol. Chem. 266, 9610-9616, 1991
 A:Title: Structure of the human type I DNA topoisomerase gene.
 A:Reference number: A40008; MUID:91236733
 A:Accession: A40008
 A:Molecule type: DNA
 A:Residues: 1-144, 'A', 146-553, 'E', 555-765 <KUN>
 A:Cross-references: GB:M60688; GB:M60689; GB:M60690; GB:M60691; GB:M60692; GB:M60693;
 M60704; GB:M60705; GB:M60706
 R:Kunze, N.; Klein, M.; Richter, A.; Knippers, R.
 Eur. J. Biochem. 194, 323-330, 1990
 A:Title: Structural characterization of the human DNA topoisomerase I gene promoter.
 A:Reference number: S13821; MUID:91099302
 A:Accession: S13821
 A:Molecule type: DNA
 A:Residues: 1-20 <KU>
 A:Cross-references: EMBL:X52601
 R:Oddou, P.; Schmidt, U.; Knippers, R.; Richter, A.
 Eur. J. Biochem. 177, 523-529, 1988
 A:Title: Monoclonal antibodies neutralizing mammalian DNA topoisomerase I activity.
 A:Reference number: S02397; MUID:89064806
 A:Accession: S02397
 A:Molecule type: mRNA
 A:Residues: 344-765 <ODD>
 A:Cross-references: GB:M60657
 R:Tamura, H.O.; Kohchi, C.; Yamada, R.; Ikeda, T.; Koiwai, O.; Patterson, E.; Keene,
 Nucleic Acids Res. 19, 69-75, 1991
 A:Title: Molecular cloning of a cDNA of a camptothecin-resistant human DNA topoisomerase
 A:Reference number: S40643
 A:Accession: S40643
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 523-543, 573-582, 'D', 584-593 <TAM>
 R:Maul, G.G.; Jimenez, S.A.; Riggs, E.; Ziemnicka-Kotula, D.
 Proc. Natl. Acad. Sci. U.S.A. 86, 8492-8496, 1989
 A:Title: Determination of an epitope of the diffuse systemic sclerosis marker antigen
 ity in systemic sclerosis.
 A:Reference number: A34422; MUID:90046823
 A:Accession: A34422
 A:Molecule type: mRNA
 A:Residues: 657-765 <MAU>
 A:Cross-references: GB:M27913; NID:g339807; PID:g339808
 C:Comment: Type I DNA topoisomerase catalyzes the ATP-independent transient breakage
 in another, followed by rejoining. This reaction will lead to the conversion of one
 C:Genetics:
 A:Gene: GDB:TOP1
 A:Cross-references: GDB:120444; OMIM:126420
 A:Map position: 20q12-20q13.1
 A:Introns: 11/3; 20/1
 A:Note: the list of introns is incomplete
 C:Superfamily: eukaryotic type I DNA topoisomerase
 C:Keywords: DNA binding; DNA replication; heterotetramer; isomerase
 F:723/Active site: Tyr #status predicted

Query Match 57.1%; Score 40; DB 1; Length 765;
 Best Local Similarity 63.6%; Pred. No. 41;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETKEERS 14
 | | | | |
 Db 282 WRKEMTNEKN 292

RESULT 4
 C69584

aldehyde dehydrogenase aldY - Bacillus subtilis

C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
 A:Accession: C69584
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertei, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chodura, E.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinol, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, teuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:198044033
 A:Accession: C69584
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-485 <NUN>
 A:Cross-references: GB:299123; GB:A1009126; NID:g2636240; PID:ell86382; PID:g2636418
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: aldY
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 55.7%; Score 39; DB 2: Length 485;
 Best Local Similarity 54.5%; Pred. No. 38;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETKEKS 14

||| |||: |:

DB 60 WAKSTEDRKA 70

RESULT 5

TPHUN1

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 1 - human
 N:Alternate names: phosphotyrosine phosphatase 1B, long form; protein-tyrosine phosphatase
 N:Contains: protein-tyrosine-phosphatase, nonreceptor type 1, membrane-bound form; prote
 C:Species: Homo sapiens (man)
 C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 26-Feb-1999
 C:Accession: A35992; I59169; I79576; S15849; A33897; A37275
 R:Chernoff, J.; Schievella, A.R.; Jost, C.A.; Erikson, R.L.; Neel, B.G.
 Proc. Natl. Acad. Sci. U.S.A. 87, 2735-2739, 1990
 A:Title: Cloning of a cDNA for a major human protein-tyrosine-phosphatase.
 A:Reference number: A35992; MUID:90207272

A:Accession: A35992

A:Molecule type: mRNA

A:Residues: 1-435 <CHE>

A:Cross-references: GB:M31724; NID:g190741; PID:g190742

R:Brown-Shimer, S.; Johnson, K.A.; Lawrence, J.B.; Johnson, C.; Bruskin, A.; Green, N.R.

Proc. Natl. Acad. Sci. U.S.A. 87, 5148-5152, 1990

A:Title: Molecular cloning and chromosome mapping of the human gene encoding protein ph

A:Reference number: I59169; MUID:190311360

A:Accession: I59169

A:Molecule type: translated from GB/EMBL/DBJ

A:Residues: 1-435 <BRO1>

A:Cross-references: GB:M33689; NID:g190271; PID:g190272

A:Accession: I79576

A:Molecule type: translated from GB/EMBL/DBJ

A:Residues: 165-435 <BRO2>

A:Cross-references: GB:M33684; NID:g190277; PID:g190279

R:Fallen, C.J.; Lai, D.S.Y.; Chia, H.P.; Boulet, I.; Tong, P.H.

Biochem. J. 276, 315-323, 1991

A:Title: Purification and characterization of a higher-molecular-mass form of protein ph

A:Reference number: S15849; MUID:91264781

A:Accession: S15849

A:Molecule type: protein

A:Residues: 'X', 8-21:80-92, 'C', 94-99: 'GR', 123-130 <BIO>

R:Charbonneau, H.; Tonks, N.K.; Kumar, S.; Diltz, C.D.; Harrylock, M.; Cool, D.E.; Kr

Proc. Natl. Acad. Sci. U.S.A. 86, 5252-5256, 1989

A:Title: Human placenta protein-tyrosine-phosphatase: amino acid sequence and relatio

A:Reference number: A33897; MUID:89315775

A:Accession: A33897

A:Molecule type: protein

A:Residues: 1-321 <CHA>

R:Barford, D.; Flint, A.J.; Tonks, N.K.

submitted to the Brookhaven Protein Data Bank, September 1994

A:Reference number: A52879; PDB:2HNP

A:Contents: annotation; X-ray crystallography, 2.8 angstroms, residues 5-282

R:Barford, D.; Flint, A.J.; Tonks, N.K.

Science 263, 1397-1404, 1994

A:Title: Crystal structure of human protein tyrosine phosphatase 1B.

A:Reference number: A38904; MUID:94174273

A:Contents: annotation; X-ray crystallography, 2.8 angstroms

R:Flint, A.J.; Gebbink, M.F.G.B.; Franza Jr., B.R.; Hill, D.E.; Tonks, N.K.

EMBO J. 12, 1937-1946, 1993

A:Title: Multi-site phosphorylation of the protein tyrosine phosphatase, PTP1B: ident

A:Reference number: A49919; MUID:94259136

A:Contents: annotation; phosphorylation sites

C:Genetics:

A:Gene: GDB:PTPN1; PTP1B

A:Cross-references: GDB:126728; OMIM:176885

A:Map position: 20q13.1-20q13.2

A:Introns: 164/3; 234/3; 288/3; 363/2; 428/3

C:Function:

A:Description: catalyzes hydrolysis of peptidyl-phosphotyrosine to release phosphate

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 1; protein-tyrosine-pho

C:Keywords: acetylated amino end; phosphoprotein; phosphoric monoester hydrolase; tyr

F:1-435/Product: protein-tyrosine-phosphatase, nonreceptor type 1, membrane-bound for

F:1-321/Product: protein-tyrosine-phosphatase, nonreceptor type 1, membrane-bound for

F:405-425/Domain: transmembrane #status predicted <TRM>

F:1/Modified site: acetylated amino end (Met) #status experimental

F:215/Active site: Cys (phosphocysteine intermediate) #status experimental

F:221/Binding site: substrate phosphate (Arg) #status experimental

F:352/Binding site: phosphate (Ser) (covalent) (by unidentified kinase) #status exper

F:378/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status experime

F:386/Binding site: phosphate (Ser) (covalent) (by cdc2 kinase) #status experimental

Query Match 54.3%; Score 38; DB 1: Length 435;

Best Local Similarity 60.0%; Pred. No. 50;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETKEEK 13

||| |||: |:

DB 333 WVKETQEDK 342

RESULT 6

I38202

N:leupin precursor - human

C:Species: Homo sapiens (man)

C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 13-Mar-1998

C:Accession: I38202; S66675; S57522

R:Schneider, S.S.; Schick, C.; Fish, K.E.; Muller, E.; Pena, J.C.; Treter, S.D.; Hui,

Proc. Natl. Acad. Sci. U.S.A. 92, 3147-3151, 1995

A:Title: A serine proteinase inhibitor locus at 18q21.3 contains a tandem duplication

A:Reference number: I38200; MUID:95241462

A:Accession: I38202

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-390 <SCH>

A:Cross-references: EMBL:U19576; GB:U19569; NID:g852466

R:Barnes, R.C.; Worrall, D.M.

FEBS Lett. 373, 61-65, 1995

A:Title: Identification of a novel human serpin gene: cloning sequencing and expression
 A:Reference number: S66675; MUID:96013887

A:Accession: S66675
 A:Molecule type: mRNA
 A:Residues: 7-351, 'V', 353-384 <BAR>
 A:Cross-references: EMBL:X89015; NID:g887464; PID:g887465

C:Genetics:

A:Gene: GDB:P11; GDB:SCCA2
 A:Cross-references: GDB:686246; GDB:636556; OMIM:600518
 A:Map position: 18q21.3-18q21.3
 A:Introns: 55/3; 74/3; 117/3; 157/1; 204/3; 256/3
 C:Superfamily: antithrombin III
 C:Keywords: glycoprotein; serine proteinase inhibitor
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-384/Product: leupin #status predicted <MAT>
 F:65,93,170,376/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:354/Inhibitory site: Leu (unidentified proteinase) #status predicted

Query Match 54.3%; Score 38; DB 2; Length 390;
 Best Local Similarity 56.2%; Pred. NO. 45;
 Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 2 GXW-----AKETTKKEK 13

DB 184 GOWENKFKKENTKEK 199

RESULT 7

C56849
 dopamine receptor-like protein D215 - Japanese pufferfish
 C:Species: Fugu rubripes (Japanese pufferfish)
 C:Date: 27-Oct-1995 #sequence_revision 27-Oct-1995 #text_change 05-Dec-1998
 C:Accession: C56849
 R:Macrae A.D.; Brenner, S.
 Genomics 25, 436-446, 1995
 A:Title: Analysis of the dopamine receptor family in the compact genome of the puffer fish
 A:Reference number: A56849
 A:Accession: C56849
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-511 <MAC>
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: neurotransmitter receptor

Query Match 53.6%; Score 37.5; DB 2; Length 511;

Best Local Similarity 56.7%; Pred. NO. 71;
 Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 2 GXW-AKETTKKEE 12

DB 17 GTWSANETTKDE 28

RESULT 8

S57845
 protein-tyrosine kinase (EC 2.7.1.112), receptor type tie precursor - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 26-Feb-1999
 C:Accession: S57845; S32690
 R:Sato, T.N.; Qin, Y.; Kozak, C.A.; Audus, K.L.
 Proc. Natl. Acad. Sci. U.S.A. 90, 9355-9358, 1993
 A:Title: tie-1 and tie-2 define another class of putative receptor tyrosine kinase genes
 A:Reference number: S57845; MUID:94022374

A:Accession: S57845
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-1136 <SAT>
 A:Cross-references: EMBL:X71423; NID:g296575; PID:g296576
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1993
 C:Superfamily: protein-tyrosine kinase, receptor type tie; EGF homology; fibronectin tyf
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph

F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-1136/Product: protein-tyrosine kinase, receptor type tie #status predicted <MAT>
 F:36-108/Domain: immunoglobulin homology <IG1>
 F:213-253/Domain: EGF homology <EGF1>
 F:257-300/Domain: EGF homology <EGF2>
 F:304-342/Domain: EGF homology <EGF3>
 F:363-426/Domain: immunoglobulin homology <IG2>
 F:447-528/Domain: fibronectin type III repeat homology <FN3A>
 F:540-631/Domain: fibronectin type III repeat homology <FN3B>
 F:640-728/Domain: fibronectin type III repeat homology <FN3C>
 F:759-784/Domain: transmembrane #status predicted <TM>
 F:835-1112/Domain: protein kinase homology <KIN>
 F:843-851/Region: protein kinase ATP-binding motif
 F:43-106,370-424/Disulfide bonds: #status predicted
 F:84,159,501,594,707/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:868,885,977/Active site: Lys, Glu, Asp #status predicted

Query Match 52.9%; Score 37; DB 1; Length 1136;
 Best Local Similarity 54.5%; Pred. NO. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACXWAKETTKKE 11

DB 215 AGRWGQDCTKE 225

RESULT 9

RRIWV
 genome polyprotein - Marburg virus (strain Musoke)
 N:Alternate names: L protein
 N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
 C:Species: Marburg virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 29-May-1998
 C:Accession: A42450
 R:Muehlberger, E.; Sanchez, A.; Randolph, A.; Will, C.; Kiley, M.P.; Klenk, H.D.; Feld
 Virology 187, 534-547, 1992
 A:Title: The nucleotide sequence of the L gene of Marburg virus, a filovirus: homolog
 A:Reference number: A42450; MUID:92188528
 A:Accession: A42450
 A:Molecule type: genomic RNA
 A:Residues: 1-2330 <MOE>
 A:Cross-references: GB:M92834; NID:g332178; PID:g332179
 C:Genetics:
 A:Gene: L
 C:Superfamily: parainfluenza virus RNA-directed RNA polymerase
 C:Keywords: ATP; nucleotidyltransferase; RNA biosynthesis
 F:1325-1360/Domain: ATP binding #status predicted <AT1>
 F:1390-1420/Domain: ATP binding #status predicted <AT2>
 F:1560-1593/Domain: ATP binding #status predicted <AT3>
 F:1932-1961/Domain: ATP binding #status predicted <AT4>

Query Match 52.9%; Score 37; DB 1; Length 2330;
 Best Local Similarity 66.7%; Pred. NO. 4e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETTKKEE 12

DB 2042 WTAETTKDE 2050

RESULT 10

A49546
 DNA topoisomerase (EC 5.99.1.2) - Chinese hamster
 C:Species: Crictetus griseus (Chinese hamster)
 C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 08-Sep-1997
 C:Accession: A49546; S32697; S32698
 R:Taizawa, A.; Bertrand, R.; Kohlhaagen, G.; Tabuchi, A.; Jenkins, J.; Pommier, Y.
 J. Biol. Chem. 268, 25463-25468, 1993
 A:Title: Cloning of Chinese hamster DNA topoisomerase I cDNA and identification of a
 A:Reference number: A49546

A:Accession: A49546
A:Molecule type: mRNA
A:Residues: 1-767 <TA2>
A:Cross-references: GB:221625; NID:g297078; PID:g297079
A:Experimental source: DC3F cells
A:Note: this form is not camptothecin-resistant
R:ranizawa, A.; Tabuchi, A.; Bertrand, R.; Pommier, Y.
submitted to the EMBL Data Library, February 1993
A:Reference number: S32697
A:Accession: S32697
A:Molecule type: mRNA
A:Residues: 1-504, S', 506-767 <TA2>
A:Cross-references: EMBL:221624; NID:g297076; PID:g297077
A:Experimental source: DC3F/C-10 cells
A:Note: this form is camptothecin-resistant
C:Superfamily: eukaryotic type I DNA topoisomerase
C:Keywords: isomerase

Query Match 52.9%; Score 37; DB 2; Length 767;
Best Local Similarity 54.5%; Pred. NO. 1.3e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETKEKS 14
| | | | |
DB 284 WRKMTNDEKN 294

RESULT 11
I38201
squamous cell carcinoma antigen 1 - human
C:Species: Homo sapiens (man)
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 13-Mar-1998
C:Accession: I38201; G01631; J70966; J70967
R:Schneider, S.S.; Schick, C.; Fish, K.E.; Miller, E.; Pena, J.C.; Treter, S.D.; Hui, S.
Proc. Natl. Acad. Sci. U.S.A. 92, 3147-3151, 1995
A:Title: A serine proteinase inhibitor locus at 18q21.3 contains a tandem duplication of
A:Reference number: I38200; MUID:95241462
A:Accession: I38201
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-390 <SCH1>
A:Cross-references: EMBL:U19568; GB:U19558; NID:g1172085
A:Accession: I38200
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown
A:Residues: 1-117 <SCH2>
A:Cross-references: EMBL:U19562; NID:g897835; PID:g897844
R:Silverman, G.A.
submitted to the EMBL Data Library, January 1995
A:Reference number: G07968
A:Accession: G01631
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-350, G', 352-390 <SIL>
A:Cross-references: EMBL:U19556; NID:g1276435; PID:g1052869
R:Suminami, Y.; Kishi, F.; Sekiguchi, K.; Kato, H.
Biochem. Biophys. Res. Commun. 181, 51-58, 1991
A:Title: Squamous cell carcinoma antigen is a new member of the serine protease inhibitor
A:Reference number: J70966; MUID:92068241
A:Accession: J70966
A:Molecule type: mRNA
A:Residues: 1-350, G', 352-356, A', 358-390 <SUM1>
A:Cross-references: GB:S66896; NID:g239551; PID:g239552
A:Accession: J70967
A:Molecule type: protein
A:Residues: 11-21; 231-237; 240-256; 303-325 <SUN2>
C:Comment: This antigen belongs a member of the serine proteinase inhibitors family, and
C:Genetics:
A:Gene: GDB:SCC1; SCC
A:Cross-references: GDB:625364; OMIM:600517
A:Map position: 18q21.3-18q21.3
A:Introns: 55/3; 74/3; 117/3; 157/1; 204/3; 256/3

C:Superfamily: antithrombin III
C:Keywords: cysteine proteinase inhibitor; glycoprotein
F:65,93,171,376/Binding site: carbohydrate (Asn) #status predicted
F:354/Inhibitory site: Ser (cathepsin L) #status predicted

Query Match 52.9%; Score 37; DB 2; Length 390;
Best Local Similarity 56.2%; Pred. NO. 65;
Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 2 GXWA----KETTKEK 13
| | | | |
DB 184 GQWKKFNKEDTKEK 199

RESULT 12
C69421
methylinologen-reducing hydrogenase, subunit alpha (vhuA) homolog - Archaeoglobus ful
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
C:Accession: C69421
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: C69421
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-458 <KLE>
A:Cross-references: GB:AE001009; GB:AE000782; NID:g2689332; PID:g2649200; TIGR:AF1372

Query Match 52.9%; Score 37; DB 2; Length 458;
Best Local Similarity 63.6%; Pred. NO. 77;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GXWAKETKEE 12
| | | | |
DB 164 GQWAKAPTDEE 174

RESULT 13
A70460
DNA polymerase III gamma subunit - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
C:Accession: A70460
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: A70460
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-473 <AQF>
A:Cross-references: GB:AE000759; NID:g2984125; PID:g2984127; GB:AE000657
A:Experimental source: strain VF5
C:Genetics:
A:Gene: dnaX

Query Match 52.9%; Score 37; DB 2; Length 473;
Best Local Similarity 66.7%; Pred. NO. 80;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GXWAKETKEE 13
| | | | |
DB 365 GSVWAKETKEE 376

RESULT 14

I41193
Outer membrane protein eae - Escherichia coli
N:Alternate names: outer membrane protein 1
C:Species: Escherichia coli
C:Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 13-Mar-1998
C:Accession: I41193; S20027; I41191; S17357; S19838
R:Beebakhee, G.; Louie, M.; De Azavedo, J.; Brunton, J.
FEMS Microbiol. Lett. 91, 63-68, 1992
A:Title: Cloning and nucleotide sequence of the eae gene homologue from enterohaemorrhagic
A:Reference number: I41193
A:Accession: I41193
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-935 <RES>
A:Cross-references: EMBL:X60439; NID:g42155; PID:g42156
R:Yu, J.; Kaper, J.B.
Mol. Microbiol. 6, 411-417, 1992
A:Title: Cloning and characterization of the eae gene of enterohaemorrhagic Escherichia
A:Reference number: S20027; MUID:92204008
A:Accession: S20027
A:Molecule type: DNA
A:Residues: 1-220, 'D', 222-310, 'RR', 313-317, 'H', 319-641, 'S', 643-769, 771, 'GE', 774-935 <YU>
A:Cross-references: EMBL:Z11541; NID:g41333; PID:g41334
R:Louie, M.; de Azavedo, J.; Clarke, R.; Borczyk, A.; Lior, H.; Richter, M.; Brunton, J.
Epidemiol. Infect. 112, 449-461, 1994
A:Title: Sequence heterogeneity of the eae gene and detection of verotoxin-producing Esc
A:Reference number: I41191; MUID:94273785
A:Accession: I41191
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 684-769, 771, 'GE', 774-903, 'S', 905-929 <RE2>
A:Cross-references: GB:L08095; NID:g499364; PID:g538254
C:Genetics:
A:Gene: eae
C:Superfamily: Invasin
C:Keywords: membrane protein

Query Match 51.4%; Score 36; DB 1; Length 935;

Best Local Similarity 45.5%; Pred. No. 2.3e+02;

Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETKEEKS 14

| | | | | | | |

Db 895 WIKQTSSEQRS 905

RESULT 15

S10512
Phosphoserine transaminase (EC 2.6.1.52) - Salmonella gallinarum
C:Species: Salmonella gallinarum
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 05-Dec-1998
C:Accession: S10512
R:Griffin, H.G.
Nucleic Acids Res. 18, 4260, 1990
A:Title: Nucleotide sequence of the Salmonella serC gene.
A:Reference number: S10512; MUID:90332434
A:Accession: S10512
A:Molecule type: DNA
A:Residues: 1-362 <GRI>
A:Cross-references: EMBL:X53381; NID:g47093; PID:g47094
A:Experimental source: strain 9
C:Genetics:
A:Gene: serC
C:Function:
A:Pathway: serine biosynthesis
C:Superfamily: phosphoserine aminotransferase
C:Keywords: aminotransferase; phosphoprotein; pyridoxal phosphate; serine biosynthesis
F:198/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 51.4%; Score 36; DB 2; Length 362;

Best Local Similarity 53.8%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AGXWAKETKEEK 13

| | | | | | |

Db 99 AGYWAASAIKEAK 111

Search completed: November 13, 1999, 12:08:00
Job time: 2072 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:33:35 ; Search time 51.07 Seconds
(without alignments)
7.749 Million cell updates/sec

Title: US-08-913-430-7

Perfect score: 70

Sequence: 1 AGXWAKETTREKS 14

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	40	57.1	765	1	TOP1_HUMAN
2	38	54.3	760	1	METE_MYCLE
3	38	54.3	435	1	PTN1_HUMAN
4	38	54.3	390	1	SCC2_HUMAN
5	37.5	53.6	463	1	D2DR_FUGRU
6	37	52.9	1181	1	ABLL_CAEEL
7	37	52.9	390	1	SCC1_HUMAN
8	37	52.9	1136	1	TIEL_BOVIN
9	37	52.9	767	1	TOP1_CRIGR
10	36	51.4	1443	1	DPO3_MYCPN
11	36	51.4	934	1	EAE2_ECOLI
12	36	51.4	297	1	SC13_YEAST
13	36	51.4	361	1	SERC_ECOLI
14	36	51.4	362	1	SERC_SALGL
15	36	51.4	362	1	SERC_SALTY
16	36	51.4	310	1	SIAL_BOVIN
17	36	51.4	630	1	YD13_SCHPO
18	36	51.4	197	1	YGVV_ECOLI
19	36	51.4	141	1	YIC7_YEAST
20	35	50.0	556	1	ADEC_ARCFU
21	35	50.0	498	1	ARAA_BACSU
22	35	50.0	224	1	BCHM_RHOCA
23	35	50.0	398	1	DCAM_ORYSA
24	35	50.0	361	1	DCAM_TOBAC
25	35	50.0	4644	1	DIHC_RAT
26	35	50.0	715	1	ERF2_CANAL
27	35	50.0	685	1	ERF2_YEAST
28	35	50.0	690	1	HVUA_PESN
29	35	50.0	732	1	KELL_HUMAN
30	35	50.0	770	1	PBFC_ECOLI
31	35	50.0	362	1	SERC_BACCI
32	35	50.0	917	1	SLAP_THETH
33	35	50.0	1090	1	SVLC_YEAST
34	35	50.0	376	1	YBC9_YEAST
35	35	50.0	264	1	YJ3_YEAST
36	35	50.0	544	1	YRT3_CAEEL
37	34.5	49.3	301	1	LEG6_MOUSE
38	34	48.6	495	1	DHAL_ALTAL
39	34	48.6	1012	1	DPOG_PICPA
40	34	48.6	1504	1	DPO2_YEAST
41	34	48.6	523	1	E2BD_RABIT
42	34	48.6	741	1	ERF2_PICPI
43	34	48.6	429	1	G154_MOUSE

ALIGNMENTS

```
RESULT 1
TOP1_HUMAN
ID TOP1_HUMAN STANDARD; PRT; 765 AA.
AC P11387;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DNA TOPOISOMERASE I (EC 5.99.1.2).
GN TOP1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88190108.
RA D'ARPA P., WACHLIN P.S., RATHIE H. III, ROTHFIELD N.F.,
RA CLEVELAND D.W., EARNshaw W.C.;
RT "CDNA cloning of human DNA topoisomerase I: catalytic activity of a
RT 67.7-kDa carboxyl-terminal fragment.";
PROC. NATL. ACAD. SCI. U.S.A. 85:2543-2547(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91236733.
RA KUNZE N., YANG G., DOLBERG M., SUNDARP R., KNIPPERS R., RICHTER A.;
RT "Structure of the human type I DNA topoisomerase gene.";
J. BIOL. CHEM. 266:9610-9616(1991).
RN [3]
RP SEQUENCE OF 541-765 FROM N.A.
RX MEDLINE; 89288043.
RA ZHOU B.S., BASTOW K.F., CHENG Y.C.;
RT "Characterization of the 3' region of the human DNA topoisomerase I
RT gene.";
CANCER RES. 49:3922-3927(1989).
RN [4]
RP SEQUENCE OF 657-765 FROM N.A.
RX MEDLINE; 90046823.
RA MAUL G.G., JIMENEZ S.A., RIGGS E., ZIEMNICKA-KOTULA D.;
RT "Determination of an epitope of the diffuse systemic sclerosis marker
RT antigen DNA topoisomerase I: sequence similarity with retroviral
RT p30gag protein suggests a possible cause for autoimmunity in systemic
RT sclerosis.";
PROC. NATL. ACAD. SCI. U.S.A. 86:8492-8496(1989).
RN [5]
RP VARIANTS CPT-RESISTANT.
RX MEDLINE; 91187651.
RA TAMURA H., KOHCHI C., YAMADA R., IKEDA T., KOIWA O., PATTERSON E.,
RA KEENE J.D., OKADA K., KJELDSEN E., NISHIKAWA K.;
RT "Molecular cloning of a cDNA of a camptothecin-resistant human DNA
RT topoisomerase I and identification of mutation sites.";
NUCLEIC ACIDS RES. 19:69-75(1991).
RN [6]
RP VARIANT CPT-RESISTANT ALA-729.
RX MEDLINE; 93075133.
RA KUBOTA N., KANZAWA F., NISHIO K., TAKEDA Y., OHMORI T., FUJIWARA Y.,
RA TERASHIMA Y., SAJO N.;
RT "Detection of topoisomerase I gene point mutation in CPT-11 resistant
RT lung cancer cell line.";
BIOCHEM. BIOPHYS. RES. COMMUN. 188:571-577(1992).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 215-765.
RX MEDLINE; 98155246.
RA REDINO M.R., STEWART L., KUHN P., CHAMPOUX J.J., HOL W.G.J.;
RT "Crystal structures of human topoisomerase I in covalent and
RT noncovalent complexes with DNA.";
SCIENCE 279:1504-1513(1998).
RN [8]
```

Q63342 rattus norv
P34602 caenorhabdi

RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE: 90311360.
 RA BROWN-SHIMER S., JOHNSON K.A., LAWRENCE J.B., JOHNSON C.,
 RA BRUSHIN A., GREEN N.R., HILL D.E.;
 RT "Molecular cloning and chromosome mapping of the human gene encoding
 RT protein phosphotyrosyl phosphatase 1B.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:5148-5152(1990).
 RN [3]
 RP SEQUENCE OF 1-321.
 RC TISSUE=PLACENTA;
 RX MEDLINE: 89315775.
 RA CHARBONNEAU H., TONKS N.K., KUMAR S., DILTZ C.D., HARRYLOCK M.,
 RA COOL D.E., KREBS E.G., FISCHER E.H., WALSH K.A.;
 RT "Human placenta protein-tyrosine-phosphatase: amino acid sequence and
 RT relationship to a family of receptor-like proteins.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 86:5252-5256(1989).
 RN [4]
 RP PHOSPHORYLATION SITES.
 RX MEDLINE: 93259136.
 RA FLINT A.J., GEBBINK M.F.G.B., FRANZA B.R. JR., HILL D.E., TONKS N.T.;
 RT "Multi-site phosphorylation of the protein tyrosine phosphatase,
 RT PTP1B: identification of cell cycle regulated and phorbol ester
 RT stimulated sites of phosphorylation.";
 RL EMBO J. 12:1937-1946(1993).
 RN [5]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE: 92154669.
 RA FRANGIONI J.V., BEAHM P.H., SHIFRIN V., JOST C.A., NEEL B.G.;
 RT "The nontransmembrane tyrosine phosphatase PTP-1B localizes to the
 RT endoplasmic reticulum via its 35 amino acid C-terminal sequence.";
 RL CELL 68:545-560(1992).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 1-321.
 RX MEDLINE: 94174273.
 RA BARFORD D., FLINT A.J., TONKS N.K.;
 RT "Crystal structure of human protein tyrosine phosphatase 1B.";
 RL SCIENCE 263:1397-1404(1994).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-298 OF MUTANT SER-215.
 RX MEDLINE: 98054248.
 RA PUJOS Y.A., ZHAO Y., SULLIVAN M., LAWRENCE D.S., ALMO S.C.,
 RA ZHANG Z.Y.;
 RT "Identification of a second aryl phosphate-binding site in protein-
 RT tyrosine phosphatase 1B: a paradigm for inhibitor design.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 94:13420-13425(1997).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 1-285.
 RX MEDLINE: 98221181.
 RA PANNIFER A.D., FLINT A.J., TONKS N.K., BARFORD D.;
 RT "Visualization of the cysteinyl-phosphate intermediate of a protein-
 RT tyrosine phosphatase by X-ray crystallography.";
 RL J. BIOL. CHEM. 273:10454-10462(1998).
 CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: ASSOCIATED TO THE ENDOPLASMIC RETICULUM VIA
 CC ITS C-TERMINAL DOMAIN WITH ITS PHOSPHATASE DOMAIN ORIENTED TOWARDS
 CC THE CYTOPLASM.
 CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M31724; G190742;
 DR EMBL: M33689; G190272;
 DR EMBL: M33684; G190279;
 DR EMBL: M33688; G190279; JOINED.

DR EMBL: M33687; G190279; JOINED.
 DR EMBL: M33686; G190279; JOINED.
 DR EMBL: M33685; G190279; JOINED.
 DR PTR: A35992; TPHUN1.
 DR PDB: 2HNP; 20-DEC-94.
 DR PDB: 2HNO; 20-DEC-94.
 DR PDB: 1PTT; 01-AUG-96.
 DR PDB: 1PTU; 01-AUG-96.
 DR PDB: 1PTV; 01-AUG-96.
 DR PDB: 1PTY; 21-JAN-98.
 DR PDB: 1AAX; 04-MAR-98.
 DR PDB: 1A5Y; 17-JUN-98.
 DR MW: 176885;
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS00356; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
 DR PFAM; PF0102; Y_phosphatase; 1.
 KW HYDROLASE; ACETYLATION; PHOSPHORYLATION; 3D-STRUCTURE.
 FT MOD_RES 1 1
 FT DOMAIN 40 288
 FT ACT_SITE 215 215
 FT MOD_RES 352 352
 FT MOD_RES 378 378
 FT MOD_RES 386 386
 FT MOD_RES 435 AA; 49966 MW; 3F44235D CRC32;
 SQ SEQUENCE 435 AA; 49966 MW; 3F44235D CRC32;
 Query Match 54.3%; Score 38; DB 1; Length 435;
 Best Local Similarity 60.0%; Pred. No. 32;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 4 WAKETKEEK 13
 DB 333 WAKEETQEDK 342
 RESULT 4
 SCC2_HUMAN
 ID SCC2_HUMAN STANDARD; PRT; 390 AA.
 AC P48594;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE SQUAMOUS CELL CARCINOMA ANTIGEN 2 (SCCA-2) (LEUPIN).
 GN SCCA2.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95241462.
 RA SCHNEIDER S.S., SCHICK C., FISH K.E., MILLER E., PENA J.C.,
 RA TRETER S.D., HUI S.M., SILVERMAN G.A.;
 RT "A serine proteinase inhibitor locus at 18q21.3 contains a tandem
 RT duplication of the human squamous cell carcinoma antigen gene.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 92:3147-3151(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96013887.
 RA BARNES R.C., WORRALL D.M.;
 RT "Identification of a novel human serpin gene; cloning sequencing and
 RT expression of leupin.";
 RL FEBS LETT. 373:61-65(1995).
 CC -!- FUNCTION: MAY ACT AS A PROTEASE INHIBITOR TO MODULATE THE HOST
 CC IMMUNE RESPONSE AGAINST TUMOR CELLS.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- TISSUE SPECIFICITY: SQUAMOUS CELLS.
 CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

DR	EMBL: X80175; E112838; -.	
DR	GCRDB: GCR_1107; -.	
DR	PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.	
DR	PFAM: PF00001; 7tm_1; 1.	
KW	G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;	
KW	MULTIGENE FAMILY.	
FT	DOMAIN 1	35
FT	TRANSMEM	58
FT	DOMAIN 59	69
FT	TRANSMEM	70
FT	DOMAIN 96	106

CC entities requires a license agreement (see <http://www.isib.sib.ch/>)
or send an email to license@isib-sib.ch).

CC
CC
CC
DR EMBL; Z50806; G951003; -;
DR EMBL; MJ3235; G55072; -;
DR FIR; A00629; TVKW6.
DR WORMPEP; M79.1; CE18889.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00001; SH2; 1.

DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; P11362; 1FCI.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; SH2 DOMAIN; SH3 DOMAIN.
FT DOMAIN 100 173 SH3.
FT DOMAIN 179 269 SH2.
FT DOMAIN 296 547 PROTEIN KINASE.
FT NP_BIND 302 310 ATP (BY SIMILARITY).
FT BINDING 325 325 ATP (BY SIMILARITY).
FT ACT_SITE 417 417 BY SIMILARITY.
SQ SEQUENCE 1181 AA; 133576 MW; 64072BDD CRC32;

Query Match 52.9%; Score 37; DB 1; Length 1181;
Best Local Similarity 80.08; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AKETTKEEKS 14
| | | | |
DB 796 ADETTKEKS 805

RESULT 7
SCC1_HUMAN STANDARD; PRT; 390 AA.
AC P29508;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE SQUAMOUS CELL CARCINOMA ANTIGEN 1 (SCCA-1) (PROTEIN T4-A).
GN SCCA1 OR SCCA.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 92068241.
RA SUMINAMI Y., KISHI F., SEXIGUCHI K., KATO H.;
RT "Squamous cell carcinoma antigen is a new member of the serine
protease inhibitors.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 181:51-58(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95241452.
RA SCHNEIDER S.S., SCHICK C., FISH K.E., MILLER E., PENA J.C.,
RT "A serine proteinase inhibitor locus at 18q21.3 contains a tandem
duplication of the human squamous cell carcinoma antigen gene.";
RL PROC. NATL. ACAD. SCI. U.S.A. 92:3147-3151(1995).
CC -1- FUNCTION: MAY ACT AS A PROTEASE INHIBITOR TO MODULATE THE HOST
IMMUNE RESPONSE AGAINST TUMOR CELLS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: SQUAMOUS CELLS.
CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION IS CLOSELY RELATED TO CELLULAR
DIFFERENTIATION IN BOTH NORMAL AND MALIGNANT SQUAMOUS CELLS.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; S66896; G239552; -
DR EMBL; U19556; G1276436; -
DR EMBL; U19568; G1172087; -
DR EMBL; U19559; G1172087; JOINED.
DR EMBL; U19560; G1172087; JOINED.
DR EMBL; U19562; G1172087; JOINED.

DR EMBL; U19565; G1172087; JOINED.
DR EMBL; U19567; G1172087; JOINED.
DR EMBL; U19562; G897844; -
DR EMBL; U19559; G897844; JOINED.
DR EMBL; U19560; G897844; JOINED.
DR PIR; J0966; J0966.
DR MIM; 600517; -
DR PROSITE; PS00284; SERPIN; 1.
DR PFAM; PF00079; serpin; 1.
DR HSSP; P01008; 1ATH.
KW SERPIN; SERINE PROTEASE INHIBITOR.
FT ACT_SITE 354 355 REACTIVE BOND.
FT CONFLICT 357 357 A -> T (IN REF. 2).
SQ SEQUENCE 390 AA; 44534 MW; 4FF63A61 CRC32;

Query Match 52.9%; Score 37; DB 1; Length 390;
Best Local Similarity 56.2%; Pred. No. 42;
Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 2 GXWA----KETTKEEK 13
| | | | |
DB 184 QWEKKNKEDTKEEK 199

RESULT 8
TIE1_BOVIN STANDARD; PRT; 1136 AA.
ID TIE1_BOVIN
AC Q06805;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE RECEPTOR TIE-1 PRECURSOR (EC 2.7.1.112).
GN TIE1 OR TIE-1.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX SATO T.N., QIN Y., KOZAK C.A., ANDUS K.L.;
RT "Tie-1 and tie-2 define another class of putative receptor tyrosine
kinase genes expressed in early embryonic vascular system.";
RL PROC. NATL. ACAD. SCI. U.S.A. 90:9355-9358(1993).
CC -1- FUNCTION: PROBABLE PROTEIN TYROSINE-KINASE TRANSMEMBRANE RECEPTOR.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN DEVELOPING VASCULAR
ENDOTHELIAL CELLS.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE
EXTRACELLULAR DOMAINS CONTAINS 2 IG-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; X71423; G296576; -
DR PIR; S32690; S32690.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 3.

DR PFAM; PF00008; EGF; 2.
DR PFAM; PF00041; fn3; 3.
DR PFAM; PF00069; pkinaase; 1.
DR HSP; P11362; lfc1.
KW RECEPTOR; TYROSINE-PROTEIN KINASE; TRANSFERASE; SIGNAL; ATP-BINDING;
KW REPEAT; EGF-LIKE DOMAIN; TRANSMEMBRANE; IMMUNOGLOBULIN FOLD;
KW GLYCOPROTEIN; PHOSPHORYLATION; MULTIGENE FAMILY.
FT SIGNAL 1 23
FT CHAIN 24 1136 TYROSINE-PROTEIN KINASE RECEPTOR TIE-1.
FT DOMAIN 24 757 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 758 782 POTENTIAL.
FT DOMAIN 783 1136 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 43 106 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 212 254 EGF-LIKE 1.
FT DOMAIN 256 301 EGF-LIKE 2.
FT DOMAIN 303 343 EGF-LIKE 3.
FT DOMAIN 370 424 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 444 538 FIBRONECTIN TYPE-III.
FT DOMAIN 541 637 FIBRONECTIN TYPE-III.
FT DOMAIN 641 742 FIBRONECTIN TYPE-III.
FT DOMAIN 837 1116 PROTEIN KINASE.
FT NP_BIND 843 851 ATP (BY SIMILARITY).
FT BINDING 868 868 ATP (BY SIMILARITY).
FT ACT_SITE 977 977 BY SIMILARITY.
FT MOD_RES 1005 1005 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 84 84 POTENTIAL.
FT CARBOHYD 159 159 POTENTIAL.
FT CARBOHYD 501 501 POTENTIAL.
FT CARBOHYD 594 594 POTENTIAL.
FT CARBOHYD 707 707 POTENTIAL.
SQ SEQUENCE 1136 AA; 124953 MW; FFA0DFCC CRC32;

Query Match 52.9%; Score 37; DB 1; Length 1136;
Best Local Similarity 54.5%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 ACXWAKETKE 11
||| : |||
Db 215 AGRWGQDCTKE 225

RESULT 9
TOP1_CRIGR STANDARD; PRT; 767 AA.
AC Q07050;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE DNA TOPOISOMERASE I (EC 5.99.1.2).
GN TOP1 OR TOP-1.
OS CRICETULUS GRISEUS (CHINESE HAMSTER).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; CRICETINAE; CRICETULUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94064611.
RA TANIZAWA A., BERTRAND R., KOHLHAGEN G., TABUCHI A., JENKINS J.,
RA POMMIER Y.
RT "Cloning of Chinese hamster DNA topoisomerase I cDNA and
RT identification of a single point mutation responsible for
RT camptothecin resistance."
RL J. BIOL. CHEM. 268:25463-25468(1993).
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC -1- ENZYME REGULATION: SPECIFICALLY INHIBITED BY CAMPTOTHECIN (CPT), A
CC PLANT ALKALOID WITH ANTITUMOR ACTIVITY.
CC -1- SUBUNIT: MONOMER.
CC -1- EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH NEGATIVE AND
CC POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES RELAX ONLY
CC NEGATIVE SUPERCOILS.

CC -1- WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA BACKBONE BOND, IT
CC SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN WHICH A TYROSYL OXYGEN
CC IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS AT ONE END OF THE
CC ENZYME-SEVERED DNA STRAND.
CC -1- SIMILARITY: BELONGS TO EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z21624; G297077; -.
CC EMBL; Z21625; G297079; -.
CC PIR; S32697; S32697.
CC DR PROSITE; PS00176; TOPOISOMERASE_I_EUK; 1.
CC DR PFAM; PF01028; Topoisomerase_I; 1.
CC DR HSP; P11387; IA35.
CC KW ISOMERASE; TOPOISOMERASE; DNA-BINDING.
CC FT ACT_SITE 725 725 DNA CLEAVAGE (BY SIMILARITY).
CC FT VARIANT 505 505 S->G (IN CPT-RESISTANT CELL).
CC SQ SEQUENCE 767 AA; 90867 MW; 332EDD67 CRC32;
Query Match 52.9%; Score 37; DB 1; Length 767;
Best Local Similarity 54.5%; Pred. No. 83;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 4 WAKETTREKS 14
|||| : |||
Db 284 WRKMTNDKN 294
RESULT 10
DPO3_MYCPN STANDARD; PRT; 1443 AA.
AC P75080;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DNA POLYMERASE III, ALPHA CHAIN POLC-TYPE (EC 2.7.7.7) (POLIII).
GN POLC.
OS MYCOPLASMA PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 29342 / M129;
RX MEDLINE; 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERRMANN R.
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE ALPHA CHAIN HAS BOTH DNA POLYMERASE AND EXONUCLEASE ACTIVITY
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC N PYROPHOSPHATE + DNA(N).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; AE000014; G1673778; --.
DR PFAAM; PF00929; Exonuclease; 1.
KW TRANSFERASE; DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION; HYDROLASE;
KW EXONUCLEASE.
SQ SEQUENCE 1443 AA; 165101 MW; 06DC2A74 CRC32;

Query Match 51.4%; Score 36; DB 1; Length 1443;
Best Local Similarity 50.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 GWAKETKEEK 13
| | | | |
DB 277 GTWKATVKVER 288

RESULT 11
ID EAE2_ECOLI STANDARD; PRT; 934 AA.
AC P43261;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE INTIMIN (ATTACHING AND EFFACING PROTEIN) (EAE PROTEIN).
GN EAEA OR EAE.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ENTEROHEMORRHAGIC (EHEC) EDL933 / O157:H7;
RX MEDLINE; 92204008.
RA YU J., KAPER J.B.;
RT "Cloning and characterization of the eae gene of enterohaemorrhagic
Escherichia coli O157:H7";
RL MOL. MICROBIOL. 6:411-417(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ENTEROHEMORRHAGIC (EHEC) CL-8 / O157:H7;
RX MEDLINE; 92249750.
RA BEEBAKHEE G., LOUIE M., DE AZAVEDO J., BRUNTON J.;
RT "Cloning and nucleotide sequence of the eae gene homologue from
enterohaemorrhagic Escherichia coli serotype O157:H7";
RL FEMS MICROBIOL. LETT. 70:63-68(1992).
RN [3]
RP SEQUENCE OF 1-37 FROM N.A.
RC STRAIN=ENTEROPATHOGENIC (EPEC) HA1 / O157:H7;
RX MEDLINE; 96079490.
RA ZHAO S., MITCHELL S.E., MENG J., DOYLE M.P., KRESOVICH S.;
RT "Cloning and nucleotide sequence of a gene upstream of the eaeA gene
of enterohaemorrhagic Escherichia coli O157:H7";
RL FEMS MICROBIOL. LETT. 133:35-39(1995).
CC -1- FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
LESIONS ON TISSUE CULTURE CELLS.
CC -1- SUBCELLULAR LOCATION: OUTER SURFACE.
CC -1- SIMILARITY: BELONGS TO THE EAE/INVASIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11541; G41334; -
DR EMBL; G60439; G42156; -
DR EMBL; U32312; G975877; -
KW OUTER MEMBRANE.
FT CONFLICT 221 221 D -> N (IN REF. 2).
FT CONFLICT 311 312 RR -> SG (IN REF. 2).
FT CONFLICT 318 318 H -> N (IN REF. 2).
FT CONFLICT 642 642 S -> T (IN REF. 2).

FT CONFLICT 769 769 V -> VK (IN REF. 2).
FT CONFLICT 771 772 GE -> SM (IN REF. 2).
SQ SEQUENCE 934 AA; 102014 MW; 48A94D3D CRC32;

Query Match 51.4%; Score 36; DB 1; Length 934;
Best Local Similarity 45.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETKEEKS 14
| | | | |
DB 894 WIKOTSEQMS 904

RESULT 12
ID SC13_YEAST STANDARD; PRT; 297 AA.
AC Q04491;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROTEIN TRANSPORT PROTEIN SEC13.
GN SEC13 OR YLR208W OR L8167.4
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE; 93163112.
RA PRIER N.K., SALAMA N.R., SCHEKMAN R., KAISER C.A.;
RT "Cytosolic Sec13p complex is required for vesicle formation from the
endoplasmic reticulum in vitro";
RL J. CELL BIOL. 120:865-875(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU 2.,
RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J., KUCABA T.,
RA HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D.,
RA JOHNSTON L., LANGSTON Y., LATREILLE P., MARDIS E., MENEZES S.,
RA MILLER N., NHAN M., PAULEY A., PELUSO D., RIFKEN L., RILES L.,
RA TAICH S., TREVASKIS E., VAUDIN M., VIGNATI D., WILCOX L., WILSON R.,
RA WOHLDMAN P., WATERSTON R.;
RL SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 35-39; 79-83; 125-131; 278-282 AND 289-293.
RX MEDLINE; 96152656.
RA SINIOSSOGLOU S., WIMMER C., RIEGER M., DOYE V., TEKOTTE H., WEISE C.,
RA EMIG S., SEGREF A., HURT E.C.;
RT "A novel complex of nucleoporins, which includes Sec13p and a Sec13p
homolog, is essential for normal nuclear pores";
RL CELL 84:265-275(1996).
CC -1- FUNCTION: COMPONENT OF THE COPII COAT, THAT COVERS ER-DERIVED
VESICLES INVOLVED IN TRANSPORT FROM THE ENDOPLASMIC RETICULUM TO
THE GOLGI APPARATUS. COPII IS COMPOSED OF AT LEAST FIVE PROTEINS:
THE SEC23/24 COMPLEX, THE SEC13/31 COMPLEX, AND THE PROTEIN SAR1.
CC REQUIRED IN VESICLE BIOGENESIS AT A STEP BEFORE OR CONCURRENT WITH
THE RELEASE OF TRANSPORT VESICLES FROM THE ER MEMBRANE. REQUIRED
FOR GERMINATION AND/OR GROWTH AT 24 DEGREES CELSIUS. REQUIRED FOR
EFFICIENT MRNA EXPORT FROM THE NUCLEUS TO THE CYTOPLASM AND FOR
CORRECT NUCLEAR PORE BIOGENESIS.
CC -1- SUBUNIT: FORMS AN ACTIVE 700 KD LARGE COMPLEX WITH OTHER PROTEINS.
CC INTERACTS WITH NUP84, NUP85, NUP120 AND SEH1. MIGHT INTERACT WITH
SEC23 AND SAR1.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. PERIPHERALLY ASSOCIATED WITH
MEMBRANES.
CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE SEC13 FAMILY OF WD-REPEAT PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way


```

Db      98 AGYWAASAIKEAK 110
RESULT 14
SERC_SALGL STANDARD; PRT; 362 AA.
AC P17902;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHOSPHOSERINE AMINOTRANSFERASE (EC 2.6.1.52) (PSAT).
GN SERC.
OS SALMONELLA GALLINARUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9;
RX MEDLINE; 90332434.
RA GRIFFIN H.G.;
RT "Nucleotide sequence of the salmonella serC gene.";
RL NUCLEIC ACIDS RES. 18:4260-4260(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=9;
RX MEDLINE; 91259030.
RA GRIFFIN H.G., GRIFFIN A.M.;
RT "Cloning and DNA sequence analysis of the serC-aroA operon from
RT Salmonella gallinarum: evolutionary relationships between the
RT prokaryotic and eukaryotic aroA-encoded enzymes.";
RL J. GEN. MICROBIOL. 137:113-121(1991).
CC -!- CATALYTIC ACTIVITY: O-PHOSPHO-L-SERINE + 2 OXOGLUTARATE =
CC 3-PHOSPHONOOXYPYRUVATE + L-GLUTAMATE.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- PATHWAY: REQUIRED BOTH IN MAJOR PHOSPHORYLATED PATHWAY OF SERINE
CC BIOSYNTHESIS AND IN THE BIOSYNTHESIS OF PYRIDOXINE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X53381; G47094; -.
DR EMBL; M62801; G154365; -.
DR PIR; S10512; S10512.
DR PROSITE; PS00595; AA_TRANSFER_CLASS_5; 1.
DR PFAM; PF00266; aminotran_5; 1.
DR HSP; P23721; 1BJN.
KW SERINE BIOSYNTHESIS; PYRIDOXINE BIOSYNTHESIS; TRANSFERASE;
KW AMINOTRANSFERASE; PYRIDOXAL PHOSPHATE.
FT BINDING 198 198 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 362 AA; 39855 MW; 549E57E1 CRC32;

Query Match 51.4%; Score 36; DB 1; Length 362;
Best Local Similarity 53.8%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AGXWAKETTKEEK 13
Db 99 AGYWAASAIKEAK 111

RESULT 15
SERC_SALTY STANDARD; PRT; 362 AA.
AC P17902;
DT 01-NOV-1990 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHOSPHOSERINE AMINOTRANSFERASE (EC 2.6.1.52) (PSAT).
GN SERC.
OS SALMONELLA TYPHIMURUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RA MOUSLIM C.; FLORES A.; CANO D.A.; CASADESUS J.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: O-PHOSPHO-L-SERINE + 2 OXOGLUTARATE =
CC 3-PHOSPHONOOXYPYRUVATE + L-GLUTAMATE.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- PATHWAY: REQUIRED BOTH IN MAJOR PHOSPHORYLATED PATHWAY OF SERINE
CC BIOSYNTHESIS AND IN THE BIOSYNTHESIS OF PYRIDOXINE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; Y10355; E290666; -.
DR STYGENE; SG77777; SERC.
DR PROSITE; PS00595; AA_TRANSFER_CLASS_5; 1.
DR PFAM; PF00266; aminotran_5; 1.
DR HSP; P23721; 1BJN.
KW SERINE BIOSYNTHESIS; PYRIDOXINE BIOSYNTHESIS; TRANSFERASE;
KW AMINOTRANSFERASE; PYRIDOXAL PHOSPHATE.
FT BINDING 198 198 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 362 AA; 39823 MW; CE5B8A0D CRC32;

Query Match 51.4%; Score 36; DB 1; Length 362;
Best Local Similarity 53.8%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AGXWAKETTKEEK 13
Db 99 AGYWAASAIKEAK 111

Search completed: November 13, 1999, 10:33:37
Job time: 5187 sec

```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:28 ; Search time 139.86 Seconds
(without alignments)
6.161 Million cell updates/sec

Title: US-08-913-430-7
Perfect score: 70
Sequence: 1 AGXWAKETKEKS 14

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database : SPTREMBL10:**
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	61.4	766	13	P79994 gallus gall
2	41	58.6	26926	4	Q10466
3	41	58.6	153	10	O49649 arabidopsis
4	41	58.6	526	10	O80728 arabidopsis
5	40	57.1	761	4	Q12855 homo sapien
6	39	55.7	485	2	P94358 bacillus su
7	39	55.7	1802	6	Q28633 oryctolagus
8	39	55.7	2000	6	O97791 oryctolagus
9	37	52.9	458	1	O28899 archaeoglob
10	37	52.9	473	2	O67707 aquifex aeo
11	37	52.9	328	3	O74776 schizosacch
12	37	52.9	700	3	O74936 yarrowia li
13	37	52.9	4981	5	O77372 plasmodium
14	37	52.9	660	5	Q23673 caenorhabdi
15	37	52.9	367	10	P93032 arabidopsis
16	37	52.9	367	10	O23007 arabidopsis
17	36	51.4	934	2	O85627 escherichia
18	36	51.4	195	2	Q92KL0
19	36	51.4	767	2	O92HY4
20	36	51.4	435	2	Q923C0
21	36	51.4	1030	4	O14974 homo sapien
22	36	51.4	450	5	Q21549 caenorhabdi
23	36	51.4	976	11	Q10728 rattus norv
24	36	51.4	1004	13	Q90623 gallus gall
25	36	51.4	963	13	Q90624
26	35	50.0	496	2	O05184 bacillus su
27	35	50.0	98	2	O69086 escherichia
28	35	50.0	1878	2	Q59986 streptococ
29	35	50.0	176	2	P96854 mycobacteri

ALIGNMENTS

RESULT 1

P79994
ID P79994 PRELIMINARY; PRT; 766 AA.
AC P79994;
DT 01-MAY-1997 (TREMREL. 03, Created)
DT 01-MAY-1997 (TREMREL. 03, Last sequence update)
DT 01-MAY-1999 (TREMREL. 10, Last annotation update)
DE DNA TOPOISOMERASE I.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RA SUKA N.;
RT "Association of DNA topoisomerase I and nucleolin with the
W-heterochrom atic body of chicken."
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB000407; BAA19101.1;
DR PFAM; PF01028; Topoisomerase I; 1.
DR PROSITE; PS00176; TOPOISOMERASE_I_EUK; 1.
KW isomerase.
SQ SEQUENCE 766 AA; 90796 MW; 1C0B5112 CRC32;

Query Match 61.4%; Score 43; DB 13; Length 766;
Best Local Similarity 72.7%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETKEKS 14

Db 283 WRKMTSEKS 293

RESULT 2

Q10466
ID Q10466 PRELIMINARY; PRT; 26926 AA.
AC Q10466;
DT 01-NOV-1996 (TREMREL. 01, Created)
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
DE TITIN, HEART ISOFORM N2-B (EC 2.7.1.-) (CONNECTIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE; 96026330
RA LABEIT S.; KOLMER B.;
RT "Titins: giant proteins in charge of muscle ultrastructure and elasticity."
RL Science 270:293-296(1995).

RN RP SEQUENCE OF 22277-25376 FROM N.A.
RX MEDLINE; 92258380.
RA LABEIT S., GAUTEL M., LAKEY A., TRINICK J.;
RT "Towards a molecular understanding of titin."
RL EMBO J. 11:1711-1716(1992).
[3]
RN RP SEQUENCE OF 1976-2014 FROM N.A.
RA LABEIT S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
[4]
RN RP CHARACTERIZATION.
RX MEDLINE; 95331314.
RA GAUTEL M., CASTIGLIONE-MORELLI M.A., PFUHL M., MOTTA A., PASTORE A.;
RT "A calmodulin-binding sequence in the C-terminus of human cardiac
titin kinase."
RL Eur. J. Biochem. 230:752-759(1995).
[5]
RN RP FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
DISTRIBUTION.
CC DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY ONE TISSUE.
CC THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM N2-B.
CC -1- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
KINASES.
CC -1- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS 112 IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN
TYPE III-LIKE DOMAINS.
CC EMBL; X64698; CAA45939.1; -
CC EMBL; X83270; CAA58243.1; -
CC EMBL; X64597; CAA45938.1; -
CC EMBL; X90568; CAA62188.1; -
CC EMBL; X64699; CAA45940.1; -
CC PFAM; PF00041; fn3; 132.
CC PFAM; PF00047; Ig; 58.
CC PFAM; PF00069; pkinase; 1.
KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
KW Serine/threonine-protein kinase; Alternative splicing; Repeat;
KW Immunoglobulin fold; Phosphorylation.
FT DOMAIN 1370 1389 4 X 5 AA TANDEM REPEATS OF R-N-S-P-A.
FT DOMAIN 4429 4614 GLU/LYS/PRO/VAL-RICH.
FT DOMAIN 24731 25070 CATALYTIC.
FT DOMAIN 25030 25056 CALMODULIN-BINDING.
FT MOD_RES 1372 1372 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 1377 1377 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 1382 1382 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 1387 1387 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 26171 26171 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 26178 26178 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 26184 26184 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 26190 26190 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 22277 22277 T -> P (IN REF. 2).
FT CONFLICT 22449 22449 E -> G (IN REF. 2).
FT CONFLICT 22454 22454 T -> Q (IN REF. 2).
FT CONFLICT 23324 23324 S -> L (IN REF. 2).
SQ SEQUENCE 26926 AA; 2993428 MW; 0214E3A3 CRC32;

Query Match 58.6%; Score 41; DB 4; Length 26926;
Best Local Similarity 70.08; Pred. No. 1.4e+03;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 WAKETKEEK 13
| | | | |
Db 1995 WTKELTEEEK 2004

RESULT 3
O49649 PRELIMINARY; PRT; 153 AA.
ID O49649

O49649;
AC 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE MYB-LIKE PROTEIN.
GN T12H17.70.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA BEVAN M., HILBERT H., BRAUN M., HOLZER E., BRANDT A., DUESTERHOEF A.,
RA BANCROFT I., MEWES H.W., MAYER K., SCHUELLER C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN RP SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021635; CAA16553.1; -
DR PFAM; PF00249; myb_DNA-binding; 1.
SQ SEQUENCE 153 AA; 17556 MW; A799C771 CRC32;

Query Match 58.6%; Score 41; DB 10; Length 153;
Best Local Similarity 63.6%; Pred. No. 8.3;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 WAKETKEEKS 14
| : | | | | |
Db 91 WSDETTKDEAS 101

RESULT 4
O80728 PRELIMINARY; PRT; 526 AA.
AC O80728;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PUTATIVE CYTOCHROME P450.
GN FL4M4.21.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV, COLUMBIA;
RA ROUNSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,
RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;
RT "Arabidopsis thaliana chromosome II BAC FL4M4 genomic sequence."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004411; AAC34227.1; -
DR PFAM; PF00067; P450; 1.
SQ SEQUENCE 526 AA; 60479 MW; 4BDBD55E CRC32;

Query Match 58.6%; Score 41; DB 10; Length 526;
Best Local Similarity 77.8%; Pred. No. 28;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 WAKETKEE 12
| | | | |
Db 185 WRKEETKEE 193

RESULT 5
Q12855 PRELIMINARY; PRT; 761 AA.
ID Q12855
AC Q12855; Q12856; O43256;

DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JAN-1999 (Tremblrel. 09, Last annotation update)
 DE DNA TOPOISOMERASE I (EC 5.99.1.2) (FRAGMENT).
 GN TOPI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 [1]
 RP SEQUENCE FROM N.A.; VARIANTS CPT-RESISTANT THR-370 AND SER-722.
 RC TISSUE-PERIPHERAL BLOOD;
 RX MEDLINE: 95188170.
 RA FUJIMORI A., HARKER W.G., KOHLHAGEN G., HOKI Y., POMMIER Y.;
 RT "Mutation at the catalytic site of topoisomerase I in CEM/C2, a human
 RT leukemia cell line resistant to camptothecin.";
 RL Cancer Res. 55:1339-1346(1995).
 [2]
 RP SEQUENCE OF 433-761 FROM N.A.
 RA BRIDGEMAN A.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE OF 537-761 FROM N.A.
 RX MEDLINE: 89288043.
 RA ZHOU B.S., BASTOW K.F., CHENG Y.C.;
 RT "Characterization of the 3' region of the human DNA topoisomerase I
 RT gene.";
 RL Cancer Res. 49:3922-3927(1989).
 CC -!- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
 CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
 CC -!- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
 CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
 CC -!- ENZYME REGULATION: SPECIFICALLY INHIBITED BY CAMPTOTHECIN (CPT), A
 CC PLANT ALKALOID WITH ANTITUMOR ACTIVITY.
 CC -!- SUBUNIT: MONOMER.
 CC -!- EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH NEGATIVE AND
 CC POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES RELAX ONLY
 CC NEGATIVE SUPERCOILS.
 CC -!- WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA BACKBONE BOND, IT
 CC SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN WHICH A TYROSYL OXYGEN
 CC IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS AT ONE END OF THE
 CC ENZYME-SEVERED DNA STRAND.
 CC -!- SIMILARITY: BELONGS TO EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.
 DR AARHUS/GHENT-2DPAGE; 610; NEPHGE.
 DR MIM; 126420; -.
 DR EMBL; U07804; AAB60379.1; -.
 DR EMBL; U07806; AAB60380.1; -.
 DR EMBL; X16479; CAA34500.1; ALT_INIT.
 DR EMBL; AL022394; CAA18536.1; -.
 DR PROSITE; PS00176; TOPOISOMERASE_I_EUK; 1.
 DR PFAM; PF01028; Topoisomerase_1; 1.
 KW Isomerase; Topoisomerase; DNA-binding; Polymorphism; 3D-structure.
 FT NON_TER 1 1
 FT DOMAIN 162 165 POLY-GLU.
 FT DOMAIN 187 193 POLY-LYS.
 FT ACT_SITE 723 723 DNA CLEAVAGE (BY SIMILARITY).
 FT VARIANT 366 366 M -> T (IN CPT-RESISTANT CELL LINE
 FT CEM/C2).
 FT VARIANT 718 718 N -> S (IN CPT-RESISTANT CELL LINE
 FT CEM/C2).
 FT SEQUENCE 761 AA; 90335 MW; C36BAF6E CRC32;

Query Match 57.1%; Score 40; DB 4; Length 761;
 Best Local Similarity 63.6%; Pred. No. 60;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETKEEKS 14
 ||| ||| |||
 Db 278 WRKMTNEKN 288

RESULT 6
 P94358

ID P94358 PRELIMINARY; PERT; 485 AA.
 AC P94358;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE HIGHLY HOMOLOGOUS TO MANY ALDEHYDE DEHYDROGENASES.
 GN ALDY.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BGSC 1A1;
 RX MEDLINE: 97124196.
 RA YOSHIDA K., SHINDO K., SANO H., SEKI S., FUJIMURA M., YANAI N.,
 RA MWA Y., FUJITA Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 RT containing the lic and cel loci, and creation of a 177 kb contig
 RT covering the gnt-sacxy region.";
 RL Microbiology 142:3113-3123(1996).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE: 98044033.
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
 RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
 RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GUISEPPI G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A.,
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., IYAI M., JONES L.,
 RA JOKIS B., KARAWATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
 RA KURIITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 RA NOONE D., O'REILLY M., OGAWA K., OGURA A., ODEGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTETELLE D., PORKOLLIK S., PRESCOTT A.M.,
 RA PRESCAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHE E., ROCHE B., ROSE M., SADAIE Y.,
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
 RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDI B.,
 RA SORIN A., TACONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
 RA VIARI A., WAMBUUT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D83026; BAA11721.1; -.
 DR EMBL; Z99123; CAB15909.1; -.
 DR PFAM; PF00171; aldedh; 1.
 SQ SEQUENCE 485 AA; 52821 MW; FE74742A CRC32;

Query Match 55.7%; Score 39; DB 2; Length 485;
 Best Local Similarity 54.5%; Pred. No. 56;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETKEEKS 14
 ||| ||| |||
 Db 60 WAKSTTEDRKA 70

```
RESULT 7
Q28633 PRELIMINARY; PRT; 1802 AA.
ID Q28633
AC Q28633
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE METHYLOXEN-REDUCING HYDROGENASE, SUBUNIT ALPHA (VHUA).
DE TITIN (FRAGMENT).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-CARDIAC;
RX MEDLINE; 96087661.
RA SEBASTYEN M.G., WOLFF J.A., GREASER M.L.;
RT "Characterization of a 5.4 kb cDNA fragment from the z-line region of
RT rabbit cardiac titin reveals phosphorylation sites for
RT proline-directed kinases."
RL J. Cell Sci. 108:3029-3037(1995).
DR EMBL; U28657; AAC48494.1; -.
DR PFAM; PF00047; 1g; 9.
FT NON_TER 1
FT NON_TER 1802
SQ SEQUENCE 1802 AA; 203991 MW; CDAAD3A4 CRC32;

Query Match 55.7%; Score 39; DB 6; Length 1802;
Best Local Similarity 70.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETKEEK 13
DB 768 WTKELTAEK 777

RESULT 8
Q97791 PRELIMINARY; PRT; 2000 AA.
ID Q97791
AC Q97791
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DE TITIN (FRAGMENT).
DE TITIN (FRAGMENT).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SOLEUS SKELETAL MUSCLE;
RA GREGORIO C.C., TROMBITAS T., KOLMERER B., STIER G., GRANZIER H.,
RA KUNKE K., SUZUKI K., OBERMAYR F., HERRMANN B., SORIMACHI H.,
RA LABEIT S.;
RT "The N terminal of titin spans the Z-Disc. Its interaction with a
RT novel 19 kDa Ligand (T-cap) is required for sarcomeric integrity."
RL J. Cell Biol. 143:1-15(1998).
DR EMBL; Y18102; CAAY7028.1; -.
FT NON_TER 2000
SQ SEQUENCE 2000 AA; 222702 MW; C6DA2109 CRC32;

Query Match 55.7%; Score 39; DB 6; Length 2000;
Best Local Similarity 70.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETKEEK 13
DB 1988 WTKELTAEK 1997

RESULT 9
Q28899 PRELIMINARY; PRT; 458 AA.
ID Q28899
AC Q28899
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE METHYLOXEN-REDUCING HYDROGENASE, SUBUNIT ALPHA (VHUA).
GN AF1372.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KATNE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic,
RT sulphate-reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL; AE001009; AAB89872.1; -.
DR TIGR; AF1372; -.
DR PFAM; PF00374; Nifese_Hases; 2.
KW Hypothetical protein.
SQ SEQUENCE 458 AA; 50586 MW; 87AAC74C CRC32;

Query Match 52.9%; Score 37; DB 1; Length 458;
Best Local Similarity 63.6%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GXWAKETKEE 12
DB 164 GGWAKAPTDEE 174

RESULT 10
Q67707 PRELIMINARY; PRT; 473 AA.
ID Q67707
AC Q67707
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE DNA POLYMERASE III GAMMA SUBUNIT.
GN DNAX.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
```

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE000759; AAC07663.1; -
 SQ SEQUENCE 473 AA; 54332 MW; A73016E1 CRC32;

Query Match 52.9%; Score 37; DB 2; Length 473;
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GXWAKETKEK 13
 | ||| ||||
 Db 365 GSVVKEKKEK 376

RESULT 11

O74776 PRELIMINARY; PRT; 328 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE CONSERVED ATP-GTP BINDING PROTEIN.

GN SPBC25B2.04C.

OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-972H-;

RA LYNE M., RAVANDREAM M.A., BARRELL B.G., LEILAURE V., GALIBERT F;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL031853; CAA21262.1; -
 SQ SEQUENCE 328 AA; 37871 MW; CB364D33 CRC32;

Query Match 52.9%; Score 37; DB 3; Length 328;
 Best Local Similarity 63.6%; Pred. No. 83;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETKEKS 14
 | ||| |||
 Db 102 WFKETSTPKS 112

RESULT 12

O74936 PRELIMINARY; PRT; 700 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE ACYL-COA OXIDASE 3 (EC 1.3.3.6).

GN AC03.

OS Yarrowia lipolytica (Candida lipolytica).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Dipodascaceae; Yarrowia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-W29;

RA NICAUD J.M.J.;

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-W29;

RA LE CLAINCHE A.;

RL Thesis (1997); Institut National Agronomique Paris-Grignon.

DR EMBL; AJ001301; CAA04661.1; -

DR PFAM; PF00441; Acyl-CoA_db; 1.

KW Oxidoreductase.

SQ SEQUENCE 700 AA; 78009 MW; 63D4EE10 CRC32;

Query Match 52.9%; Score 37; DB 3; Length 700;
 Best Local Similarity 54.5%; Pred. No. 1.7e+02;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGXWAKETKE 11
 : | | : | | |
 Db 93 SGHWKXDTKE 103

RESULT 13

O77372 PRELIMINARY; PRT; 4981 AA.

ID O77372

AC O77372

DT 01-NOV-1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE MAL3P6.23 PROTEIN.

GN MAL3P6.23.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-3D7;

RA MURPHY L., LAWSON D., BARRELL B.;

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z98551; CAB1128.1; -

SQ SEQUENCE 4981 AA; 593251 MW; 183E8965 CRC32;

Query Match 52.9%; Score 37; DB 5; Length 4981;
 Best Local Similarity 70.0%; Pred. No. 1.2e+03;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETKEK 13
 | ||| |||
 Db 2830 WKKEETKEK 2839

RESULT 14

O23673 PRELIMINARY; PRT; 660 AA.

ID O23673

AC Q23673

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)

DE ZK938.3 PROTEIN.

GN ZK938.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RA LLOYD C.;

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKES M.,

RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;

RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans".

RL Nature 368:32-38(1994).

DR EMBL; Z49913; CAA90143.1; -

DR PFAM; PF00612; IQ; 1.

SQ SEQUENCE 660 AA; 76216 MW; C58C4072 CRC32;

Query Match 52.9%; Score 37; DB 5; Length 660;
 Best Local Similarity 41.7%; Pred. No. 1.6e+02;
 Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GXWAKETKEEK 13
 | | | : | : |
 Db 460 GTWQKQSKDQK 471

RESULT 15

P93032
 ID P93032 PRELIMINARY; PRT; 367 AA.
 AC P93032;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
 DE NAD+ DEPENDENT ISOCITRATE DEHYDROGENASE SUBUNIT 2.
 GN IDH2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98187249.
 RA BEHAL R.H., OLIVER D.J.;
 RT "NAD(+)-dependent isocitrate dehydrogenase from Arabidopsis thaliana.
 RL Plant Mol. Biol. 36:691-698(1998).
 DR EMBL; U81994; AAC49965.1; -.
 DR MENDEL; 7009; ARATH;1399;1.
 DR PFAM; PF00180; isodh; 2.
 SQ SEQUENCE 367 AA; 39615 MW; F566647B CRC32;

Query Match 52.9%; Score 37; DB 10; Length 367;
 Best Local Similarity 50.0%; Pred. No. 92;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AGXWAKETKEEKS 14
 | | | : | : |
 Db 292 AGNVGKDTTEQKN 305

Search completed: November 13, 1999, 12:55:29
 Job time: 3028 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:59:49 ; Search time 104.22 Seconds
(without alignments)
3.182 Million cell updates/sec

Title: US-08-913-430-7

Perfect score: 70

Sequence: 1 AGXWAKETKEERS 14

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	97.1	14	1 W01030	Mycoplasma 52-54 k
2	40	57.1	765	1 P92275	Human topoisomerase
3	40	57.1	765	1 R75915	Variant human DNA
4	38	54.3	435	1 R13907	Human protein phos
5	38	54.3	390	1 W15242	Psoriasisatin type
6	37	52.9	390	1 R25276	SCC antigen. New D
7	37	52.9	390	1 W15241	Psoriasisatin type
8	36	51.4	935	1 R55552	HEHC eae gene prod
9	36	51.4	935	1 R55550	Amino acid sequenc
10	36	51.4	934	1 W34451	Enterohaemorrhagic
11	36	51.4	934	1 W31367	Enterohaemorrhagic
12	36	51.4	976	1 W41377	Rat protein p138.
13	36	51.4	1030	1 W41378	Human protein p164
14	36	51.4	976	1 W53571	Rat p138 protein.
15	36	51.4	1030	1 W53572	Human myosin light
16	35	50.0	3025	1 P93284	Sequence of clone
17	35	50.0	44	1 P90404	Plasmodium falcipa
18	35	50.0	690	1 R15482	Protein encoded by
19	35	50.0	661	1 R63226	Mycoplasma gallise
20	35	50.0	661	1 R79910	M.gallisepticum 66
21	35	50.0	737	1 R97232	Kell protein seque
22	35	50.0	732	1 W06402	Kell protein seque
23	35	50.0	495	1 W53949	Bacillus subtilis
24	34	48.6	495	1 R71803	Alternaria alterna
25	34	48.6	16	1 R71807	A.alternata allerg
26	34	48.6	771	1 R80495	Japanese oyster tr
27	34	48.6	312	1 W06933	Cagi locus product
28	34	48.6	2213	1 W26356	Rabbit LDL recepto
29	34	48.6	3224	1 W54235	Human Nup358 prote
30	34	48.6	394	1 W49881	Open reading frame
31	34	48.6	516	1 W69426	Human secreted pro
32	34	48.6	171	1 W98273	H. pylori GHPO 610
33	34	48.6	450	1 W30644	A. thaliana xylan
34	33	47.1	2409	1 R12609	Versican. Versican
35	33	47.1	510	1 R15354	Protein deduced fr
36	33	47.1	56	1 R70181	SPR037 signal sequ
37	33	47.1	878	1 R78608	Murine IL-3 recept
38	33	47.1	878	1 R92529	Fas sequence from
39	33	47.1	499	1 W37508	Human GSPT1-TK. No
40	33	47.1	329	1 W41790	Canine IL-12 P40 s
41	33	47.1	329	1 W41791	Canine IL-12 P40 s
42	33	47.1	62	1 W27911	Staphylococcus aur
43	33	47.1	458	1 W79948	Aspergillus phoeni

ALIGNMENTS

RESULT 1

W01030
ID W01030 standard; Peptide; 14 AA.
AC W01030;
DT 19-JAN-1997 (first entry).
DE Mycoplasma 52-54 kDa protective antigen N-terminal peptide.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumonia;
diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
FH Key Location/Qualifiers
FT misc_difference 3 /note= "undetermined amino acid"
FT PN W09628472-A1.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (OYME) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
WPI; 96-433763/43.
DR Putative protective antigens against Mycoplasma - used for the
detection, prevention or treatment of Mycoplasma infections, esp. M.
hyopneumoniae in swine
PT hyopneumoniae in swine
PS Claim 10; Page 27; 49pp; English.
CC A 52-54 kDa putative protective antigen against Mycoplasma contains
the N-terminal sequence given in W01030 and the internal CNBr
fragments given in W01031-32. The antigen was isolated from
Mycoplasma hyopneumoniae cells using antibody probes enriched with
Mycoplasma-specific antibodies. Other protective antigens were
also identified (see also W01024-29 and W01033-37). Protective
antigens and antibodies can be used in vaccines for preventing or
treating mycoplasma infections, partic. M. hyopneumoniae
infections in swine. They can also be used for diagnosis.
CC Sequence 14 AA;
SQ

Query Match 97.1%; Score 68; DB 1; Length 14;
Best Local Similarity 100.0%; Pred. No. 2e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGXWAKETKEERS 14

DB 1 AGXWAKETKEERS 14

RESULT 2

P92275
ID P92275 standard; peptide; 765 AA.
AC P92275;
DT 27-Feb-1990 (first entry)
DE Human topoisomerase I cDNA
KW Scleroderma. (human).
OS Homo sapiens
PN W08909222-A.
PD 05-OCT-1989.
PF 22-MAR-1989; U01116.
PR 23-MAR-1988; US-172159.
PA (BRIG) Brigham and Women's Hospital; (UYJO) John's Hopkins Univ.
WPI; 89-309500/42.
DR N-PSDB; N91475.
DT Cloned cDNA encoding eukaryotic topoisomerase I - useful for large scale
prodn. by recombinant methods
PS Claim 6; fig. 5; 28pp; English.
CC The cDNA of this can be spliced into DNA vectors and used to transform
hosts for high yield. This polypeptide (I) retains the ability to bind
autoantibodies, even though the prokaryotic host degrades transcribed (I)

DR WPI; 91-281485/38.

QY 2 GXW---AKETKEK 13

```

Db 184 GOWENKFKENTKEEK 199
| | | | | | | | | |
RESULT 6
ID R25276 standard; Protein; 390 AA.
AC R25276;
DT 08-JAN-1993 (first entry)
DE SCC antigen.
KW Squamous cell carcinoma; cancer; immunogen; monoclonal antibody;
OS Synthetic.
FH Key Location/Qualifiers
FT region 11..21
FT /label= 21-2=26
FT /note= "region corresp. to partially determined
FT antigen sequences"
FT 231..237
FT /label= 10
FT /note= "region corresp. to partially determined
FT antigen sequences"
FT 241..256
FT /label= 21-1
FT /note= "region corresp. to partially determined
FT antigen sequences"
FT 303..325
FT /label= 19
FT /note= "region corresp. to partially determined
FT antigen sequences"
FT 65..67
FT /note= "glycosylation site"
FT 93..95
FT /note= "glycosylation site"
FT 171..173
FT /note= "glycosylation site"
FT 376..378
FT /note= "glycosylation site"
DE4139418-A.
PN 09-JUL-1992.
PD 29-NOV-1991; 139418.
PF 30-NOV-1990; JP-330155.
PR (DAIN-) DRAINBOT KK.
PA (DAIN-) DRAINBOT KK.
PI Kato H, Sekiguchi K, Suminami Y, Takeda K;
DR WPI; 92-235415/29.
DR N-PSDB; 025817.
PT New DNA encoding squamous cell carcinoma antigen - allowing large
PT scale antigen prodn. for use as diagnostic reagent or immunogen,
PT also useful as hybridisation probe
PS Claim 2; Page 9-10, Fig 4; Zipp; German.
CC The protein relates to an antigen (Ag) associated with squamous cell
CC carcinoma (SCC). The Ag (which can be isolated from SCC of the
CC cervix uteri, pulmonary or oesophageal carcinoma, etc.) is known as a
CC diagnostic reagent for SCC but is not easily produced on a large scale
CC from human cancer cells. By expressing the DNA sequence encoding this
CC protein, in e.g. E.coli, yeast or mammalian cells, Ag can be prep'd. in
CC large quantities. Ag can also be used as an immunogen (to produce
CC monoclonal antibodies, also useful diagnostically) while the Ag-
CC encoding gene can be detected by hybridisation with this sequence.
CC Ag is homologous with serine protease inhibitors so may be useful
CC therapeutically.
SQ Sequence 390 AA;

Query Match 52.9%; Score 37; DB 1; Length 390;
Best Local Similarity 56.2%; Pred. No. 81;
Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 2 GXWA----KETTKEEK 13
| | | | | | | | | |
Db 184 GOWENKFKENTKEEK 199
| | | | | | | | | |
RESULT 8
ID R55552 standard; Protein; 935 AA.
AC R55552;
DT 24-JAN-1995 (first entry)
DE EHEC eae gene product associated with attaching and effacing activity.
KW EHEC; enterohaemorrhagic Escherichia coli; attaching activity;
KW effacing activity; O serogroups; verotoxin producing E.coli;
KW VTEC; pathogenic E. coli; detection; invasion gene; inv;
KW verisnia pseudotuberculosis; receptor binding; vaccines;
KW enteropathogenic E.coli; EPEC.
OS Escherichia coli.
PN CA2078716-A.
PD 22-MAR-1994.
PF 21-SEP-1992; 078716.
PR 21-SEP-1992; CA-078716.
PA (MOUN) MOUNT SINAI HOSPITAL CORP.
PI Brunton J, De Azavedo J, Louie M;
DR WPI; 94-184072/23.
DR N-PSDB; Q66687.

```

```

RESULT 7
ID W15241 standard; Protein; 390 AA.
AC W15241;
DT 21-JUL-1997 (first entry)
DE Psoriasis; psoriasis; cancer; cell proliferation; therapy;
KW Psoriasis; psoriasis; cancer; cell proliferation; therapy;
KW antagonist.
OS Homo sapiens.
FH Key Location/Qualifiers
FT active_site 357
FT /note= "Thr-357 may correspond to the P'2 or P'3
FT position of reactive site of homologous
FT serpins"
WO9714425-A1.
PN 24-APR-1997.
PD 16-OCT-1996; U16599.
PF 17-OCT-1995; US-005679.
PR 20-MAR-1996; US-013755.
PA (GHO) GEN HOSPITAL CORP.
PA (SHIS) SHISEIDO CO LTD.
PI Bacin PC, Goerlinck PF, Hibino T, Takahashi T;
DR WPI; 97-244853/22.
DR N-PSDB; T66948.
PT Modulating cell proliferation or apoptosis - by modulating activity
PT of psoriastatin type I and II genes, e.g. using (ant)agonists,
PT useful for treatment of cancer or psoriasis
PS Disclosure; Page 44-45; 63pp; English.
CC Psoriastatin type I (W15241) is a 43 kDa polypeptide that is
CC expressed in psoriatic tissue, e.g. psoriatic epidermis, at least 2
CC and pref. at least 5-10 times more abundantly than in normal
CC tissue. It is a cross-class inhibitor, inhibits cathepsin L less
CC efficiently than does squamous cell carcinoma-antigen (SCC-A), does
CC not inhibit cathepsin B or cathepsin H, is active at pH 5.0, and is
CC secreted. Its amino acid sequence, deduced from an isolated cDNA
CC clone (T66948), shows 98% homology to SCC-A. A related sequence,
CC psoriastatin II (W15242), has also been identified. These
CC polypeptides can be used to identify psoriastatin antagonists useful
CC in the treatment of cell proliferation disorders, e.g. cancer and
CC psoriasis.
SQ Sequence 390 AA;

Query Match 52.9%; Score 37; DB 1; Length 390;
Best Local Similarity 56.2%; Pred. No. 81;
Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 2 GXWA----KETTKEEK 13
| | | | | | | | | |
Db 184 GOWENKFKENTKEEK 199
| | | | | | | | | |
RESULT 8
ID R55552 standard; Protein; 935 AA.
AC R55552;
DT 24-JAN-1995 (first entry)
DE EHEC eae gene product associated with attaching and effacing activity.
KW EHEC; enterohaemorrhagic Escherichia coli; attaching activity;
KW effacing activity; O serogroups; verotoxin producing E.coli;
KW VTEC; pathogenic E. coli; detection; invasion gene; inv;
KW verisnia pseudotuberculosis; receptor binding; vaccines;
KW enteropathogenic E.coli; EPEC.
OS Escherichia coli.
PN CA2078716-A.
PD 22-MAR-1994.
PF 21-SEP-1992; 078716.
PR 21-SEP-1992; CA-078716.
PA (MOUN) MOUNT SINAI HOSPITAL CORP.
PI Brunton J, De Azavedo J, Louie M;
DR WPI; 94-184072/23.
DR N-PSDB; Q66687.

```


PT DNA and protein associated with enterohaemorrhagic E.coli
 PT activity - useful for detecting pathogenic verotoxin-producing
 PT E.coli
 PS Claim 2: Page 37-46: 65pp: English.
 CC A chromosomal eae gene associated with attaching and effacing
 CC activity in enterohaemorrhagic E.coli (EHEC) (Q66687) has been
 CC identified. The amino acid sequence is translated in all 3 ORFs in
 CC the specification, but only the longest is given here. It was
 CC identified from the EHEC serotype O157:H7 strain CL-8 (the most
 CC important of the group of E.coli which produce verotoxins (VTEC) and
 CC is a major public health concern), using a Sali-StuI Ikb fragment as
 CC a probe derived from the eae gene of enteropathogenic E.coli (EPEC)
 CC using primers Q66688-89. Both the EHEC and EPEC sequences show
 CC similarity to the Yersinia pseudotuberculosis invasin gene and it is
 CC concluded that the C-terminal end of the sequences are associated
 CC with receptor binding. The O serogroup was probed with fragments of
 CC the EHEC eae gene by hybridisation methods or by amplification using
 CC PCR primers designed from the gene (Q66688-97). It was shown that
 CC it is possible to detect only pathogenic strains of the O157 VTEC.
 CC The invention provides a monoclonal antibody for detecting the strains,
 CC probes to detect its presence in humans, animals and food sources,
 CC and also vaccines that are useful in effecting immunity against
 CC diseases, eg haemorrhagic colitis and haemolytic uraemic syndrome.
 SQ Sequence 935 AA;

Query Match 51.4%; Score 36; DB 1; Length 935;
 Best Local Similarity 45.5%; Pred. No. 2.9e+02;
 Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETTKEEKS 14
 | | : | : | : |
 Db 895 WIKQTSSEQRS 905

RESULT 9

R55550
 ID R55550 standard; protein; 935 AA.
 AC R55550;
 DT 13-FEB-1995 (first entry)
 DE Amino acid sequence of 4'', acylase - macrolide antibiotic.
 KW macrolide antibiotic; 4'', acylase; acyl B2 gene; high expression;
 KW 4'' acylase.
 OS Streptomyces thermotolerans.
 PN J06121677-A.
 PD 06-MAY-1994.
 PF 23-JAN-1992; 048998.
 PR 23-JAN-1992; JP-048998.
 PA (SAOC) MERCIAN CORP.
 DR WPI: 94-185917/23.
 DR N-PSDB: Q66684.
 PT High expression method of 4'', acylase gene of macrolide
 PT antibiotic - by introducing Streptomyces species acyl B2 gene
 PT into microorganism contg. 4'', acylase gene
 PS Disclosure; Page 11-14; 32pp; Japanese.
 CC The acylase activity of a 4'', macrolide antibiotic is highly expressed
 CC by introducing a DNA fragment of the acyl B2 gene (see Q66684, which
 CC encodes the amino acid sequence shown here) from a Streptomyces
 CC thermotolerans microorganism to a microorganism containing the 4''
 CC acylase gene.
 SQ Sequence 935 AA;

Query Match 51.4%; Score 36; DB 1; Length 935;
 Best Local Similarity 45.5%; Pred. No. 2.9e+02;
 Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETTKEEKS 14
 | | : | : | : |
 Db 895 WIKQTSSEQRS 905

RESULT 10

W34451
 ID W34451 standard; Protein; 934 AA.
 AC W34451;
 DT 08-MAY-1998 (first entry)
 DE Enterohaemorrhagic E. coli strain 993 eae gene product.
 KW Enterohaemorrhagic EHEC; strain 993; outer membrane protein;
 KW intimin; eae; attaching/effacing; A/E; lesion; immunisation;
 KW vaccination; genetic immunisation; targeted delivery; antibody;
 KW epithelium; passive immunisation; diagnosis.
 OS Escherichia coli.
 FH Key Location/Qualifiers
 FT Misc_difference 2 /note= "encoded by AGT"
 FT W09740161-A1.
 PN 30-OCT-1997.
 PD 18-APR-1997; U05832.
 PF 22-APR-1996; US-015936.
 PR 19-APR-1996; US-015657.
 PA (JACK-) JACKSON FOUND ADVANCEMENT MILITARY MED.
 PI McKee ML, OBrien AD, Wachtel MR;
 DR WPI: 97-535845/49.
 DR N-PSDB: T99905.
 PT Enriched or purified intimin protein and analogues, particularly
 PT coupled to histidine tag - useful for vaccination against bacteria,
 PT e.g. enterohaemorrhagic E. coli, and as carriers for other antigens
 PT or drugs for targeting the intestinal epithelium
 PS Example 1; Fig 2; 108pp; English.
 CC The present sequence is the enterohaemorrhagic Escherichia
 CC coli (EHEC) strain 993 94 kDa outer membrane protein (intimin),
 CC designated eae, which is required for the bacterium to produce
 CC attaching/effacing (A/E) lesions in tissue culture.
 CC An enriched or purified protein comprising intimin can be used in
 CC humans or animals to induce a protective immune response against
 CC intimin expressing bacteria, and against antigens (Ag) conjugated
 CC to the protein. The use of the protein for genetic immunisation is
 CC also described. Conjugates of the protein with an Ag and/or drug
 CC can be used for targeted delivery of these materials to the
 CC epithelium. Anti-intimin antibodies can be used for passive
 CC immunisation and diagnosis.
 SQ Sequence 934 AA;

Query Match 51.4%; Score 36; DB 1; Length 934;
 Best Local Similarity 45.5%; Pred. No. 2.9e+02;
 Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETTKEEKS 14
 | | : | : | : |
 Db 894 WIKQTSSEQRS 904

RESULT 11

W31367
 ID W31367 standard; Protein; 934 AA.
 AC W31367;
 DT 11-MAY-1998 (first entry)
 DE Enterohaemorrhagic E. coli (EHEC) intimin.
 KW Enterohaemorrhagic Escherichia coli; EHEC; intimin; eae gene;
 KW adherence protein; antigen; transgenic plant; diarrhoea,
 KW haemorrhagic colitis; vaccine.
 OS Escherichia coli strain 933.
 PN W09740177-A1.
 PD 30-OCT-1997.
 PF 18-APR-1997; U05831.
 PR 22-APR-1996; US-015938.
 PR 19-APR-1996; US-015657.
 PA (JACK-) JACKSON FOUND ADVANCEMENT MILITARY MED.
 PI McKee ML, OBrien AD, Stewart CN, Wachtel MR;
 DR WPI: 97-535858/49.
 PT DNA for expression of heterologous intimin in a plant - useful to
 PT stimulate immune response against enterohaemorrhagic Escherichia
 PT coli adherence protein in animals and/or humans
 PS Example 1; Fig 2; 139pp; English.

CC This is the predicted protein sequence of the enterohaemorrhagic
 CC *Escherichia coli* (EHEC) adherence protein intimin. The invention
 CC relates to a method of stimulating an immune response in an
 CC animal, including humans, by transforming a plant with a vector
 CC encoding intimin, an intimin-like protein or a portion thereof, and
 CC administering the plant to the patient. The plant may be a monocot
 CC or a dicot and is more preferably alfalfa, carrot, canola, tobacco,
 CC banana or potato. Feeding the plant to animals and/or humans
 CC elicits the production of antibodies that protect the animals
 CC and/or humans against EHEC colonisation and infection by pathogens
 CC expressing intimin-like proteins. The intimin may also be produced
 CC as a fusion protein with one or more other antigens, thereby
 CC providing protection against pathogens that express the antigen(s).
 SQ Sequence 934 AA;

Query Match 51.4%; Score 36; DB 1; Length 934;
 Best Local Similarity 45.5%; Pred. No. 2.9e+02;
 Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 4 WAKETKEEKS 14
 | : : : : :
 Db 894 WIKOTSSSEORS 904

RESULT 12

W41377

ID W41377 standard; Protein; 976 AA.
 AC W41377;
 DT 01-JUN-1998 (first entry)
 DE Rat protein p138.
 KW Rat; protein p138; Rho protein; myosin combined subunit;
 KW combining activity.
 OS Rattus sp.
 PN J10057074-A.
 PD 03-MAR-1998.
 PF 11-JUL-1996; 201325.
 PR 05-JAN-1996; JP-017151.
 PR 20-NOV-1995; JP-325130.
 PR 05-DEC-1995; JP-344605.
 PA (KIRI) KIRIN BREWERY KK.
 DR WPI; 98-210406/19.
 DR N-PSDB; V17239.

PT New physiologically active protein p138 - useful for, e.g. screening
 PT for gene therapy for treatment of disease related to Rho protein
 PS Claim 13; Page 22-26; 36pp; Japanese.
 CC This sequence represents the rat protein p138, which is a protein of the
 CC invention. The protein has the combining activity of an active type Rho
 CC protein. The protein can also contain a phosphorylated site. The DNA
 CC sequence encoding the protein can be used to screen for a genetic drug
 CC useful for treating a disease related to Rho protein. The protein or a
 CC peptide can be used for screening for an inhibitor that inhibits the
 CC combination of active Rho protein and myosin combined subunit.
 SQ Sequence 976 AA;

Query Match 51.4%; Score 36; DB 1; Length 976;
 Best Local Similarity 50.0%; Pred. No. 3.1e+02;
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 4 WAKETKEEK 13
 | : : : : :
 Db 558 WAEDSTEREK 567

RESULT 13

W41378

ID W41378 standard; Protein; 1030 AA.
 AC W41378;
 DT 01-JUN-1998 (first entry)
 DE Human protein p164.
 KW Human; protein p164; Rho protein; myosin combined subunit;
 KW combining activity.

OS Homo sapiens.
 PN J10057074-A.
 PD 03-MAR-1998.
 PF 11-JUL-1996; 201325.
 PR 05-JAN-1996; JP-017151.
 PR 20-NOV-1995; JP-325130.
 PR 05-DEC-1995; JP-344605.
 PA (KIRI) KIRIN BREWERY KK.
 DR WPI; 98-210406/19.
 DR N-PSDB; V17260.

PT New physiologically active protein p138 - useful for, e.g. screening
 PT for gene therapy for treatment of disease related to Rho protein
 PS Claim 15; Page 28-32; 36pp; Japanese.
 CC This sequence represents the human protein p164, and is a protein of the
 CC invention. The protein has the combining activity of an active type Rho
 CC protein. The protein can also contain a phosphorylated site. The DNA
 CC sequence encoding the protein can be used to screen for a genetic drug
 CC useful for treating a disease related to Rho protein. The protein or a
 CC peptide can be used for screening for an inhibitor that inhibits the
 CC combination of active Rho protein and myosin combined subunit.
 SQ Sequence 1030 AA;

Query Match 51.4%; Score 36; DB 1; Length 1030;
 Best Local Similarity 50.0%; Pred. No. 3.2e+02;
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 4 WAKETKEEK 13
 | : : : : :
 Db 614 WAEDSTEREK 623

RESULT 14

W53571

ID W53571 standard; Protein; 976 AA.
 AC W53571;
 DT 07-JUL-1998 (first entry)
 DE Rat p138 protein.
 KW Human; myosin light chain; bound subunit; Rho type protein;
 KW phosphatase promoter; tumour diagnosis; rat; p138.
 OS Rattus rattus.
 PN J10077298-A.
 PD 24-MAR-1998.
 PF 28-FEB-1997; 061847.
 PR 11-JUL-1996; JP-201325.
 PA (KIRI) KIRIN BREWERY KK.
 DR WPI; 98-264599/24.
 DR N-PSDB; V22251.
 PT Protein from human myosin light chain gene - useful for tumour
 PT diagnosis

PS Example 2; Pages 23-27; 47pp; Japanese.
 CC The present sequence was used in the preparation of a human myosin
 CC light chain bound subunit, which binds Rho type protein, promotes
 CC myosin light chain phosphatase catalytic subunit activity, has its
 CC gene positioned at human chromosome 12q15-21.2 and has a molecular
 CC weight of 130 to 135 kDa by SDS-PAGE. The gene encoding the protein
 CC can be used for the diagnosis of tumours.
 SQ Sequence 976 AA;

Query Match 51.4%; Score 36; DB 1; Length 976;
 Best Local Similarity 50.0%; Pred. No. 3.1e+02;
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 4 WAKETKEEK 13
 | : : : : :
 Db 558 WAEDSTEREK 567

RESULT 15

W53572

ID W53572 standard; Protein; 1030 AA.
 AC W53572;

DT 07-JUL-1998 (first entry)
 DE Human myosin light chain bound subunit.
 KW Human; myosin light chain; bound subunit; Rho type protein;
 KW phosphatase promoter; tumour diagnosis.
 OS Homo sapiens.
 PN J10077299-A.
 PD 24-MAR-1998.
 PF 28-FEB-1997; 061847.
 PR 11-JUL-1996; JP-201325.
 PA (KIRI) KIRIN BREWERY KK.
 DR WPI: 98-264599/24.
 DR N-PSDB; V22252.
 PT Protein from human myosin light chain gene - useful for tumour
 PT diagnosis
 PS Claim 2: Pages 30-35; 47pp; Japanese.
 CC The present sequence is a human myosin light chain bound
 CC subunit, which binds Rho type protein, promotes myosin light chain
 CC phosphatase catalytic subunit activity, has its gene positioned at
 CC human chromosome 12q15-21.2 and has a molecular weight of 130 to
 CC 135 kDa by SDS-PAGE. The gene encoding the protein can be used for
 CC the diagnosis of tumours.
 SQ Sequence 1030 AA;

Query Match 51.4%; Score 36; DB 1; Length 1030;
 Best Local Similarity 50.0%; Pred. No. 3.2e+02;
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 4 WAKETKEK 13
 DB 614 WAEDSTEKEK 623

Search completed: November 13, 1999, 18:59:51
 Job time: 130 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:29 ; Search time 64.87 Seconds
(without alignments)
2.466 Million cell updates/sec

Title: US-08-913-430-7

Perfect score: 70

Sequence: 1 AGXWAKETKREKS 14

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/PCUS9_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	40	57.1	633	1	US-08-458-477A-5	Sequence 5, Appli
2	40	57.1	765	2	US-08-663-112-2	Sequence 2, Appli
3	38	54.3	385	1	US-08-036-210-5	Sequence 5, Appli
4	37	54.3	385	2	US-08-449-609-5	Sequence 5, Appli
5	37	52.9	390	2	US-08-568-147B-2	Sequence 2, Appli
6	36	51.4	275	1	US-08-409-452-2	Sequence 2, Appli
7	35	50.0	661	2	US-08-525-742-2	Sequence 2, Appli
8	34	48.6	2213	1	US-08-727-034-3	Sequence 3, Appli
9	34	48.6	770	1	US-08-525-654A-1	Sequence 1, Appli
10	34	48.6	771	1	US-08-525-654A-3	Sequence 3, Appli
11	34	48.6	3224	2	US-08-705-660-34	Sequence 34, Appli
12	34	48.6	1727	2	US-08-477-451-10	Sequence 10, Appli
13	34	48.6	312	2	US-08-477-451-29	Sequence 29, Appli
14	33	47.1	499	2	US-08-820-170A-40	Sequence 40, Appli
15	33	47.1	56	2	US-08-245-511-45	Sequence 45, Appli
16	32.5	46.4	367	2	US-08-515-251A-4	Sequence 4, Appli
17	32	45.7	2843	1	US-07-741-940-2	Sequence 2, Appli
18	32	45.7	2842	1	US-07-741-940-7	Sequence 7, Appli
19	32	45.7	2843	1	US-08-289-548A-2	Sequence 2, Appli
20	32	45.7	2842	1	US-08-289-548A-7	Sequence 7, Appli
21	32	45.7	482	1	US-08-528-255A-1	Sequence 1, Appli
22	32	45.7	2843	1	US-08-452-654-2	Sequence 2, Appli
23	32	45.7	2842	1	US-08-452-654-7	Sequence 7, Appli
24	32	45.7	575	1	US-08-348-920-1	Sequence 1, Appli
25	32	45.7	575	1	US-08-348-920-2	Sequence 2, Appli
26	32	45.7	804	2	US-08-785-428-2	Sequence 43, Appli
27	32	45.7	482	2	US-08-717-365-1	Sequence 46, Appli
28	32	45.7	462	2	US-08-471-033-2	Sequence 50, Appli
29	32	45.7	1346	2	US-08-471-033-23	Sequence 2, Appli
30	32	45.7	410	2	US-08-471-033-40	Sequence 7, Appli
31	32	45.7	410	2	US-08-471-033-43	Sequence 2, Appli
32	32	45.7	449	2	US-08-471-033-46	Sequence 43, Appli
33	32	45.7	1338	2	US-08-471-033-50	Sequence 46, Appli
34	32	45.7	2843	2	US-08-452-655B-2	Sequence 2, Appli
35	32	45.7	2843	2	US-08-452-655B-7	Sequence 7, Appli
36	32	45.7	462	2	US-08-471-044-2	Sequence 2, Appli
37	32	45.7	1346	2	US-08-471-044-23	Sequence 23, Appli
38	32	45.7	410	2	US-08-471-044-40	Sequence 40, Appli
39	32	45.7	410	2	US-08-471-044-43	Sequence 43, Appli

Sequence 46, Appli
Sequence 50, Appli
Sequence 2, Appli
Sequence 23, Appli
Sequence 40, Appli
Sequence 43, Appli

ALIGNMENTS

RESULT 1
US-08-458-477A-5
; Sequence 5, Application US/08458477A
; Patent No. 5723311
; GENERAL INFORMATION:
; APPLICANT: WEI, ET AL.
; TITLE OF INVENTION: Human Topoisomerase I
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,477A
; FILING DATE: June 2, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05701
; FILING DATE: 18 MAY 94
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-309 (PFI18P1)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 633 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-458-477A-5

Query Match 57.1% Score 40; DB 1; Length 633;
Best Local Similarity 63.6% Pred. No. 29;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 WAKETKREKS 14
Db 150 WRKMTNEEKN 160

RESULT 2
US-08-663-112-2
; Sequence 2, Application US/08663112
; Patent No. 5849503
; GENERAL INFORMATION:
; APPLICANT: WAGATSUMA, Masako
; APPLICANT: KURITA, No. 58495031ko
; TITLE OF INVENTION: MUTANT PROTEINS OF HUMAN DNA

;; TITLE OF INVENTION: TOPOISOMERASE I
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;; ADDRESSEE: Dunner L.L.P.
;; STREET: 1300 I Street, N.W.
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005-3315
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/663,112
;; FILING DATE: 26-NOV-1996
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Einaudi, Carolyn P.
;; REGISTRATION NUMBER: 32,220
;; REFERENCE/DOCKET NUMBER: 06609.1488-00000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-408-4000
;; TELEFAX: 202-408-4400
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 765 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-663-112-2

Query Match 57.1%; Score 40; DB 2; Length 765;
Best Local Similarity 63.6%; Pred. NO. 35;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETTKEES 14
| | | | | | | |
Db 282 WRKMTNEEKN 292

RESULT 3
US-08-036-210-5
; Sequence 5, Application US/08036210
; Patent No. 5385233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,210
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie

;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 7683-025
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212-790-9090
;; TELEFAX: 212-869-8864/9741
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 385 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: unknown
;; MOLECULE TYPE: protein
US-08-036-210-5

Query Match 54.3%; Score 38; DB 1; Length 385;
Best Local Similarity 60.0%; Pred. NO. 36;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 WAKETTKEEK 13
| | | | | | | |
Db 333 WKKEETQEDK 342

RESULT 4
US-08-449-609-5
; Sequence 5, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 385 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-449-609-5

Query Match 54.3%; Score 38; DB 2; Length 385;
Best Local Similarity 60.0%; Pred. No. 36;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 WAKETKEEK 13
| | | | | | | | | |
Db 333 WKETQEDK 342

RESULT 5
US-08-568-147B-2
; Sequence 2, Application US/08568147B
; Patent No. 5783422
; GENERAL INFORMATION:
; APPLICANT: Suminami, Yoshinori
; APPLICANT: Kato, Hiroshi
; APPLICANT: Sekiguchi, Kiyoshi
; APPLICANT: Takeda, Katsumichi
; TITLE OF INVENTION: DNA FRAGMENT CODING FOR SOUMAMOUS CELL
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,147B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 800,952
; FILING DATE: 02-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8425
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 390 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-568-147B-2

Query Match 52.9%; Score 37; DB 2; Length 390;
Best Local Similarity 56.2%; Pred. No. 54;
Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

Qy 2 GXWA----KETTKEEK 13
| | | | | | | | | |
Db 184 GWEKKFNKEDTKEEK 199

RESULT 6
US-08-409-452-2
; Sequence 2, Application US/08409452
; Patent No. 5747293
; GENERAL INFORMATION:
; APPLICANT: DOUGAN, Gordon
; APPLICANT: FRANKEL, Gad

; TITLE OF INVENTION: Antibodies to Intimin-like
; TITLE OF INVENTION: Proteins of E. coli
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE and DORR
; STREET: 1455 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DISTRICT OF COLUMBIA
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,452
; FILING DATE: 23-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Hollie L.
; REGISTRATION NUMBER: 31,321
; REFERENCE/DOCKET NUMBER: 102286.319
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 942-8400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-409-452-2

Query Match 51.4%; Score 36; DB 1; Length 275;
Best Local Similarity 45.5%; Pred. No. 55;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 4 WAKETKEEKS 14
| | | | | | | | | |
Db 235 WIKOTSSEQRS 245

RESULT 7
US-08-525-742-2
; Sequence 2, Application US/08525742
; Patent No. 5871742
; GENERAL INFORMATION:
; APPLICANT: Saito, Shuji
; APPLICANT: Ohkawa, Setsuko
; APPLICANT: Sasaki, Sakiko
; APPLICANT: Ohsawa, Ikuroh
; APPLICANT: Funato, Hiroo
; APPLICANT: Iritani, Yoshikazu
; APPLICANT: Aoyama, Shigemi
; APPLICANT: Takahashi, Kiyohito

; TITLE OF INVENTION: NEW POLYPEPTIDE, DNA ENCODING THE
; TITLE OF INVENTION: POLYPEPTIDE, RECOMBINANT VECTOR BEARING THE DNA AND
; TITLE OF INVENTION: RECOMBINANT VIRUS UTILIZING THE RECOMBINANT VECTOR AS WELL
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARMSTRONG, WESTERMAN, HATTORI, MCLELAND &
; ADDRESSEE: NAUGHTON
; STREET: 1725 K Street, Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/525.742
APPLICATION NUMBER: 25-SEP-1995
FILING DATE: 25-SEP-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 05-074139
FILING DATE: 31-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 05-245625
FILING DATE: 30-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP94/00541
FILING DATE: 31-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Mclelland, Le-Nhung
REGISTRATION NUMBER: 31,541
REFERENCE/DOCKET NUMBER: 950811
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-659-2930
TELEFAX: 202-8870357
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 661 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-742-2

Query Match 50.0%; Score 35; DB 2; Length 661;
Best Local Similarity 77.8%; Pred. No. 1.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 6 KETTREKS 14
||| ||||
Db 150 KETLKEKT 158

RESULT 8
US-08-727-034-3
Sequence 3, Application US/08727034
Patent No. 5665872
GENERAL INFORMATION:
APPLICANT: SAITO, YASHUSHI
APPLICANT: IWASAKI, AKIO
APPLICANT: ARAI, KOICHI
APPLICANT: YAMAZAKI, HIROYUKI
TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND
TITLE OF INVENTION: THE GENE CODING THEREFOR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/727.034
FILING DATE: 08-OCT-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 261440/1995
FILING DATE: 09-OCT-1995
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 102451/1996
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 80-079-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2213 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-727-034-3

Query Match 48.6%; Score 34; DB 1; Length 2213;
Best Local Similarity 60.0%; Pred. No. 9.6e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 WAKETKEEK 13
||| :|||
Db 1783 WAFDTHQEK 1792

RESULT 9
US-08-525-654A-1
Sequence 1, Application US/08525654A
Patent No. 5736356
GENERAL INFORMATION:
APPLICANT: SANO, KOICHIRO
APPLICANT: KUMAZAWA, YOSHIYUKI
APPLICANT: YASEUDA, HISASHI
APPLICANT: SEGURO, KATSUYA
APPLICANT: MOTOKI, MASAO
TITLE OF INVENTION: TRANSGLUTAMINASE ORIGINATED FROM
TITLE OF INVENTION: CRASSOSTREA GIGAS
NUMBER OF SEQUENCES: 150
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525.654A
FILING DATE: 28-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6/8283
FILING DATE: 28-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7/3876
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-760-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 770 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Crassostrea gigas
; US-08-525-654A-1

Query Match 48.6%; Score 34; DB 1; Length 770;
Best Local Similarity 54.5%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGXWAKETKE 11
||| ||| |
Db 305 AGRWTKPKD 315

RESULT 10
US-08-525-654A-3
; Sequence 3, Application US/08525654A
; Patent No. 5736356
; GENERAL INFORMATION:

; APPLICANT: SANO, KOICHIRO
; APPLICANT: KUMAZAWA, YOSHIYUKI
; APPLICANT: YASEUDA, HISASHI
; APPLICANT: SEGURO, KATSUYA
; APPLICANT: MOTOKI, MASAO
; TITLE OF INVENTION: TRANSLUTAMINASE ORIGINATED FROM
; TITLE OF INVENTION: CRASSOSTREA GIGAS
; NUMBER OF SEQUENCES: 150
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525.654A
; FILING DATE: 28-SEP-1995
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6/8283
; FILING DATE: 28-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7/3876
; FILING DATE: 13-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-760-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 771 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Crassostrea gigas
; US-08-525-654A-3

Query Match 48.6%; Score 34; DB 1; Length 771;
Best Local Similarity 54.5%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGXWAKETKE 11
||| ||| |
Db 306 AGRWTKPKD 316

RESULT 11
US-08-705-660-34
; Sequence 34, Application US/08705660
; Patent No. 5858683
; GENERAL INFORMATION:
; APPLICANT: KEESEE, SUSAN
; APPLICANT: OBAR, ROBERT
; APPLICANT: WU, YING-JYE
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL CANCER
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Testa, Hurwitz & Thibault
; STREET: 125 High St.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/705.660
; FILING DATE:
; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:
; NAME: GREENHALGH, DUNCAN A
; REGISTRATION NUMBER: 38,678
; REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3224 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-705-660-34

Query Match 48.6%; Score 34; DB 2; Length 3224;
Best Local Similarity 54.5%; Pred. No. 1.4e+03;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETKEKS 14
||| ||| |
Db 2714 WEKKPTVEKA 2724

RESULT 12
US-08-477-451-10
; Sequence 10, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street

; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94508-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1727 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-451-10

Query Match 48.6%; Score 34; DB 2; Length 1727;
Best Local Similarity 87.5%; Pred. No. 7.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 6 KETTKEEK 13
Db 335 KEITKEEK 342

RESULT 13
US-08-477-451-29
; Sequence 29, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94508-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 312 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-477-451-29

Query Match 48.6%; Score 34; DB 2; Length 312;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 6 KETTKEEK 13
Db 61 KEITKEEK 68

RESULT 14
US-08-820-170A-40
; Sequence 40, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 499 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-820-170A-40

Query Match 47.1%; Score 33; DB 2; Length 499;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 4 WAKETKEEK 13
Db 125 WALDTNQEER 134

RESULT 15
US-08-245-511-45
; Sequence 45, Application US/08245511
; Patent No. 5928900
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine

```

; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: CELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/245,511
; FILING DATE: 18-MAY-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,541
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-069 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
;
US-08-245-511-45

```

```

Query Match          47.1%; Score 33; DB 2; Length 56;
Best Local Similarity 77.8%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AKETKEEK 13
   ||| ||||
Db 36 AKEAIKEEK 44

```

Search completed: November 13, 1999, 10:56:30
Job time: 1359 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:35 ; Search time 251.81 Seconds
(without alignments)
3.520 Million cell updates/sec

Title: US-08-913-430-7
Perfect score: 70
Sequence: 1 AGXWAKETTREKS 14
Scoring table: BLOSUM62
Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA.*
1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	97.1	14	13	US-08-913-430-7
2	43	61.4	766	1	PCT-US99-00559-6
3	43	61.4	766	16	US-09-226-246-6
4	43	61.4	766	23	US-09-326-246-6
5	40	57.1	765	1	PCT-US99-00559-2
6	40	57.1	711	11	US-08-738-662A-57
7	40	57.1	633	14	US-09-033-153-5
8	40	57.1	765	16	US-09-226-246-2
9	40	57.1	305	16	US-09-270-767-45505
10	40	57.1	633	17	US-09-325-430-5
11	40	57.1	765	23	US-09-226-246-2
12	38	54.3	315	11	US-08-731-566-4
13	38	54.3	391	16	US-09-221-107-110
14	38	54.3	400	16	US-09-221-107-112
15	38	54.3	391	16	US-09-285-479-110
16	38	54.3	400	16	US-09-285-479-112
17	38	54.3	875	17	US-09-328-352-4884
18	38	54.3	210	23	US-09-417-507-31771
19	38	54.3	119	23	US-09-417-507-33718

20	38	54.3	47	24	US-60-160-203-4076	Sequence 4076, Ap
21	37	52.9	390	3	US-08-099-259-2	Sequence 2, Appli
22	37	52.9	390	9	US-08-568-147A-2	Sequence 2, Appli
23	37	52.9	390	11	US-08-731-566-2	Sequence 2, Appli
24	37	52.9	366	13	US-08-928-213-59	Sequence 59, Appl
25	37	52.9	219	13	US-08-928-213-61	Sequence 61, Appl
26	37	52.9	437	16	US-09-248-796-18035	Sequence 18035, A
27	37	52.9	101	17	US-09-328-352-5841	Sequence 5841, Ap
28	37	52.9	437	19	US-60-096-409-18035	Sequence 18035, A
29	37	52.9	496	23	US-09-417-507-36857	Sequence 36857, A
30	36.5	52.1	1005	23	US-09-417-507-36326	Sequence 36326, A
31	36	51.4	935	2	US-07-947-128A-2	Sequence 2, Appli
32	36	51.4	935	4	US-08-149-108B-2	Sequence 2, Appli
33	36	51.4	195	11	US-08-761-066-406	Sequence 406, App
34	36	51.4	774	12	US-08-827-356-2826	Sequence 2826, Ap
35	36	51.4	934	12	US-08-837-459-19	Sequence 19, Appl
36	36	51.4	934	12	US-08-840-466-19	Sequence 19, Appl
37	36	51.4	275	14	US-09-054-141-2	Sequence 2, Appli
38	36	51.4	650	15	US-09-107-532-5521	Sequence 5521, Ap
39	36	51.4	247	16	US-09-248-796-14248	Sequence 14248, A
40	36	51.4	189	17	US-09-328-352-6690	Sequence 6690, Ap
41	36	51.4	247	19	US-60-096-409-14248	Sequence 14248, A
42	36	51.4	518	23	US-09-417-507-37968	Sequence 37968, A
43	35	50.0	1861	11	US-08-790-912-4	Sequence 4, Appli
44	35	50.0	496	13	US-08-926-842B-12	Sequence 12, Appl
45	35	50.0	129	17	US-09-328-352-4959	Sequence 4959, Ap

ALIGNMENTS

RESULT 1
US-08-913-430-7
; Sequence 7, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0 - beta
; SEQ ID NO 7
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: (3)
; OTHER INFORMATION: Undetermined
US-08-913-430-7

Query Match 97.1%; Score 68; DB 13; Length 14;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGXWAKETTREKS 14
Db 1 AGXWAKETTREKS 14
RESULT 2
PCT-US99-00559-6
; Sequence 6, Application PC/TUS9900559
; GENERAL INFORMATION:
; APPLICANT: Hall, William W.

```
; APPLICANT: Song, Elizabeth S.
; TITLE OF INVENTION: USE OF TOPOISOMERASE TO ENHANCE RT-BASED MOLECULAR
; FILE REFERENCE: 8842-015-228
; CURRENT APPLICATION NUMBER: PCT/US99/00559
; CURRENT FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Avian
PCT-US99-00559-6

Query Match      61.4%; Score 43; DB 1; Length 766;
Best Local Similarity 72.7%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  4 WAKETKEEKS 14
    | | | | |
Db  283 WRKMTSEKS 293

RESULT 3
US-09-226-246-6
; Sequence 6, Application US/09226246
; GENERAL INFORMATION:
; APPLICANT: Hall, William W.
; APPLICANT: Song, Elizabeth S.
; TITLE OF INVENTION: USE OF TOPOISOMERASE TO ENHANCE RT-BASED MOLECULAR
; FILE REFERENCE: 8842-010
; CURRENT APPLICATION NUMBER: US/09/226,246
; CURRENT FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Avian
US-09-226-246-6

Query Match      61.4%; Score 43; DB 16; Length 766;
Best Local Similarity 72.7%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  4 WAKETKEEKS 14
    | | | | |
Db  283 WRKMTSEKS 293

RESULT 4
US-09-226-246-6
; Sequence 6, Application US/09226246
; GENERAL INFORMATION:
; APPLICANT: Hall, William W.
; APPLICANT: Song, Elizabeth S.
; TITLE OF INVENTION: USE OF TOPOISOMERASE TO ENHANCE RT-BASED MOLECULAR
; FILE REFERENCE: 8842-010
; CURRENT APPLICATION NUMBER: US/09/226,246
; CURRENT FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Avian
US-09-226-246-6
```

```
Query Match      61.4%; Score 43; DB 23; Length 766;
Best Local Similarity 72.7%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  4 WAKETKEEKS 14
    | | | | |
Db  283 WRKMTSEKS 293

RESULT 5
PCT-US99-00559-2
; Sequence 2, Application PC/TUS9900559
; GENERAL INFORMATION:
; APPLICANT: Hall, William W.
; APPLICANT: Song, Elizabeth S.
; TITLE OF INVENTION: USE OF TOPOISOMERASE TO ENHANCE RT-BASED MOLECULAR
; FILE REFERENCE: 8842-015-228
; CURRENT APPLICATION NUMBER: PCT/US99/00559
; CURRENT FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 765
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US99-00559-2

Query Match      57.1%; Score 40; DB 1; Length 765;
Best Local Similarity 63.6%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY  4 WAKETKEEKS 14
    | | | | |
Db  282 WRKMTNEERN 292

RESULT 6
US-08-738-662A-57
; Sequence 57, Application US/08738662A
; GENERAL INFORMATION:
; APPLICANT: Foster, Jennifer M.
; APPLICANT: Giles, Kellie M.
; APPLICANT: Taylor, Alison
; APPLICANT: McGonigal, Thomas P.
; APPLICANT: Sathy, Aparna V.
; TITLE OF INVENTION: Candida Topoisomerase I Gene
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,662A
; FILING DATE: 25-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5818.US.01
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: 847-938-3137
TELEFAX: 847-938-2623
TELEX:
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: None
US-08-738-662A-57

Query Match 57.1%; Score 40; DB 11; Length 711;
Best Local Similarity 63.6%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETTKEEKS 14
| | | | |
Db 228 WRKMTNEERN 238

RESULT 7
US-09-033-153-5

; Sequence 5, Application US/09033153
; GENERAL INFORMATION:
; APPLICANT: WEI, ET AL.
; TITLE OF INVENTION: Human Topoisomerase I
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068

COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,153
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/458,477
; FILING DATE: June 2, 1995
; APPLICATION NUMBER: PCT/US94/05701
; FILING DATE: 18 MAY 94
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-309 (PFI18P1)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 633 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-09-033-153-5

Query Match 57.1%; Score 40; DB 14; Length 633;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETTKEEKS 14

Db 150 WRKMTNEERN 160
| | | | |

RESULT 8
US-09-226-246-2
; Sequence 2, Application US/09226246
; GENERAL INFORMATION:
; APPLICANT: Hall, William W.
; APPLICANT: Song, Elizabeth S.
; TITLE OF INVENTION: USE OF TOPOISOMERASE TO ENHANCE RT-BASED MOLECULAR
; TITLE OF INVENTION: BIOLOGY TECHNIQUES AND RT-BASED GENE DELIVERY
; FILE REFERENCE: 8842-010
; CURRENT APPLICATION NUMBER: US/09/226,246
; CURRENT FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 765
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-226-246-2

Query Match 57.1%; Score 40; DB 16; Length 765;
Best Local Similarity 63.6%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETTKEEKS 14
| | | | |
Db 282 WRKMTNEERN 292

RESULT 9
US-09-270-767-45505

; Sequence 45505, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45505
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: xaa means any amino acid
US-09-270-767-45505

Query Match 57.1%; Score 40; DB 16; Length 305;
Best Local Similarity 63.6%; Pred. No. 56;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 WAKETTKEEKS 14
| | | | |
Db 176 WAKRTVANKKS 186

RESULT 10
US-09-325-430-5

; Sequence 5, Application US/09325430
; GENERAL INFORMATION:
; APPLICANT: WEI, ET AL.
; TITLE OF INVENTION: Human Topoisomerase I
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD

CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/325.430
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/458.477
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-309 (PF118P1)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 633 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-09-325-430-5

Query Match 57.1%; Score 40; DB 17; Length 633;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 WAKETKEEKS 14
| | | | |
Db 150 WRKEMTNEKN 160

RESULT 11
US-09-226-246-2
Sequence 2, Application US/09226246
GENERAL INFORMATION:
APPLICANT: Hall, William W.
TITLE OF INVENTION: USE OF TOPOISOMERASE TO ENHANCE RT-BASED MOLECULAR
DELIVERY
FILE REFERENCE: 8842-010
CURRENT APPLICATION NUMBER: US/09/226.246
CURRENT FILING DATE: 1999-01-07
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 765
TYPE: PRT
ORGANISM: Homo sapiens
US-09-226-246-2

Query Match 57.1%; Score 40; DB 23; Length 765;
Best Local Similarity 63.6%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 WAKETKEEKS 14
| | | | |
Db 282 WRKEMTNEKN 292

RESULT 12
US-08-731-566-4

Sequence 4, Application US/08731566
GENERAL INFORMATION:
APPLICANT: Toshihiko Hibino et al.
TITLE OF INVENTION: TWO NOVEL GENES FROM PSORIATIC EPIDERMIS:
PSORIASTATIN TYPE I AND PSORIASTATIN TYPE II
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1775
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/731.566
FILING DATE: 16-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/013,755
FILING DATE: 20-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/005,679
FILING DATE: 17-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Louis Myers
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-040CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 315 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-731-566-4

Query Match 54.3%; Score 38; DB 11; Length 315;
Best Local Similarity 56.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

Qy 2 GXW----AKETKEEK 13
| | | | |
Db 109 GQWENKFKKENTKEEK 124

RESULT 13
US-09-221-107-110
Sequence 110, Application US/09221107
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
FILE REFERENCE: 21021.455C2
CURRENT APPLICATION NUMBER: US/09/221.107
CURRENT FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 161
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 110
LENGTH: 391
TYPE: PRT
ORGANISM: Homo sapiens
US-09-221-107-110

Query Match 54.3%; Score 38; DB 16; Length 391;
Best Local Similarity 56.2%; Pred. No. 1.6e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

Oy 2 GXW-----AKETTKEEK 13
 | | | | |
 Db 185 QWDREFKKEEK 200

RESULT 14
 US-09-221-107-112
 ; Sequence 112, Application US/09221107
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
 ; FILE REFERENCE: 210121.455C2
 ; CURRENT APPLICATION NUMBER: US/09/221.107
 ; CURRENT FILING DATE: 1998-12-22
 ; NUMBER OF SEQ ID NOS: 161
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 112
 ; LENGTH: 400
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-221-107-112

Query Match 54.3%; Score 38; DB 16; Length 400;
 Best Local Similarity 56.2%; Pred. No. 1.6e+02;
 Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

Oy 2 GXW-----AKETTKEEK 13
 | | | | |
 Db 194 QWDREFKKEEK 209

RESULT 15
 US-09-285-479-110
 ; Sequence 110, Application US/09285479
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF
 ; TITLE OF INVENTION: LUNG CANCER
 ; FILE REFERENCE: 210121.455C3
 ; CURRENT APPLICATION NUMBER: US/09/285.479
 ; CURRENT FILING DATE: 1999-04-02
 ; NUMBER OF SEQ ID NOS: 172
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 110
 ; LENGTH: 391
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 US-09-285-479-110

Query Match 54.3%; Score 38; DB 16; Length 391;
 Best Local Similarity 56.2%; Pred. No. 1.6e+02;
 Matches 9; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

Oy 2 GXW-----AKETTKEEK 13
 | | | | |
 Db 185 QWDREFKKEEK 200

Search completed: November 13, 1999, 05:08:36
 Job time: 10862 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:59:51 ; Search time 104.22 Seconds
(without alignments)
2.273 Million cell updates/sec

Title: US-08-913-430-8
Perfect score: 55
Sequence: 1 AWTADGTVN 10

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	100.0	10	1 W01031	Mycoplasma 52-54 k
2	41	74.5	21	1 W01032	Mycoplasma 52-54 k
3	36	65.5	544	1 W28866	Rat Brain Neurogly
4	35	63.6	782	1 R15625	Cellulase AE-1. Ce
5	35	63.6	428	1 R1372	Murine bone morpho
6	35	63.6	438	1 R86905	Murine BMP-9. BMP
7	35	63.6	572	1 W02256	Mouse L-glutamate
8	35	63.6	574	1 W26598	Human glutamate tr
9	35	63.6	574	1 W58553	Human excitatory a
10	35	63.6	574	1 W83923	Human excitatory a
11	35	63.6	640	1 Y13394	Amino acid sequenc
12	34	61.8	2273	1 R98811	Erysiphe graminis
13	34	61.8	2089	1 W08333	Cyclothea cryptic
14	34	61.8	401	1 W99396	S. antibioticus ole
15	34	61.8	541	1 Y01648	Cytochrome P450 en
16	34	61.8	541	1 Y01647	Cytochrome P450 de
17	33	60.0	377	1 R06692	Feline infectious
18	33	60.0	377	1 R10934	Feline Infectious
19	33	60.0	315	1 R13159	2-ketoaldehyde red
20	33	60.0	382	1 R39224	Nucleocapsid prote
21	33	60.0	1313	1 R36781	A3 maize ACCase. D
22	33	60.0	2240	1 R67819	Acetyl CoA carboxy
23	33	60.0	2254	1 R76949	ACCase. Plant acet
24	33	60.0	377	1 R78144	FIPV N protein. Re
25	33	60.0	785	1 R84617	Oilseed rape acety
26	33	60.0	2325	1 W05590	Maize acetyl CoA c
27	33	60.0	2260	1 W05212	Wheat acetyl-CoA c
28	33	60.0	2257	1 W05209	Wheat acetyl-CoA c
29	33	60.0	2325	1 W56736	Maize ACCase enzym
30	33	60.0	2172	1 W70409	Wheat acetyl-CoA c
31	33	60.0	2257	1 W70407	Triticum aestivum
32	33	60.0	140	1 W67831	Kidney injury asso
33	33	60.0	372	1 W67857	Human secreted pro
34	33	60.0	617	1 Y06785	M. grisea PTH11 ge
35	32	58.2	221	1 R20180	Sequence of the 23
36	32	58.2	465	1 R63201	Wheat germ ACCase
37	32	58.2	116	1 R66307	Human immunoglobul
38	32	58.2	517	1 R84619	Wheat acetyl-Coenz
39	32	58.2	2199	1 R94562	Human cytotactin.
40	32	58.2	300	1 W7626	Staphylococcus aur
41	32	58.2	603	1 W83214	Human h-NUMB. New
42	32	58.2	609	1 W83215	Human h-NUMB-R. Ne
43	31	56.4	470	1 R04945	Equine hemagglutin

ALIGNMENTS

RESULT 1

W01031
ID W01031 standard; Peptide; 10 AA.
AC W01031:
DT 19-JAN-1997 (first entry)
DE Mycoplasma 52-54 kDa protective antigen fragment CNBR F1.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumoniae;
diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
PN W09628472-Al.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (UYME) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
DR WPI: 96-433763/43.
PT Putative protective antigens against Mycoplasma - used for the
detection, prevention or treatment of Mycoplasma infections, esp. M.
hyopneumoniae in swine
PS Claim 11; Page 27; 43pp; English.
CC A 52-54 kDa putative protective antigen against Mycoplasma contains
the N-terminal sequence given in W01030 and the internal CNBR
fragments given in W01031-32. The antigen was isolated from
Mycoplasma hyopneumoniae cells using antibody probes enriched with
CC Mycoplasma-specific antibodies. Other protective antigens were
also identified (see also W01024-29 and W01033-37). Protective
CC antigens and antibodies can be used in vaccines for preventing or
treating mycoplasma infections, partic. M. hyopneumoniae
CC infections in swine. They can also be used for diagnosis.
Sequence 10 AA;

Query Match 100.0%; Score 55; DB 1; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.00028; Mismatches 0; Indels 0; Gaps 0;

QY 1 AWTADGTVN 10

|||||

Db 1 AWTADGTVN 10

RESULT 2

W01032
ID W01032 standard; Peptide; 21 AA.
AC W01032:
DT 19-JAN-1997 (first entry)
DE Mycoplasma 52-54 kDa protective antigen fragment CNBR F2.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumoniae;
diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
PN W09628472-Al.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (UYME) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
DR WPI: 96-433763/43.
PT Putative protective antigens against Mycoplasma - used for the
detection, prevention or treatment of Mycoplasma infections, esp. M.
hyopneumoniae in swine
PS Claim 11; Page 27; 43pp; English.
CC A 52-54 kDa putative protective antigen against Mycoplasma contains
the N-terminal sequence given in W01030 and the internal CNBR
fragments given in W01031-32. The antigen was isolated from
Mycoplasma hyopneumoniae cells using antibody probes enriched with
CC Mycoplasma-specific antibodies. Other protective antigens were

CC also identified (see also W01024-29 and W01033-37). Protective
 CC antigens and antibodies can be used in vaccines for preventing or
 CC treating mycoplasma infections, partic. M. hyopneumoniae
 CC infections in swine. They can also be used for diagnosis.
 SQ Sequence 21 AA;

Query Match 74.5%; Score 41; DB 1; Length 21;
 Best Local Similarity 90.0%; Pred. No. 0.21;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AWTADGTN 10
 I | | | | |
 Db 1 AWTADGTN 10

RESULT 3
 W28866 standard; Protein; 544 AA.
 ID W28866;
 AC W28866;
 DT 21-NOV-1997 (first entry)
 DE Rat Brain Neuroglycan C glycoprotein.
 KW NGC: chondroitin sulphate; Sprague-Dawley rat; neuraminidase;
 KW proteoglycan; central nervous system; O-glycosidase; N-glycosidase.
 OS Rattus sp.
 FH Key Location/Qualifiers
 FT peptide 1..30
 FT /label= signal
 FT modified_site 37..38
 FT /label= glycosaminoglycan_attachment_site
 FT modified_site 48..49
 FT /label= glycosaminoglycan_attachment_site
 FT modified_site 79..80
 FT /label= glycosaminoglycan_attachment_site
 FT modified_site 122..124
 FT /label= glycosaminoglycan_attachment_site
 FT modified_site 141..142
 FT /label= glycosaminoglycan_attachment_site
 FT modified_site 156..157
 FT /label= glycosaminoglycan_attachment_site
 FT modified_site 187..188
 FT /label= glycosaminoglycan_attachment_site
 FT modified_site 236..237
 FT /label= glycosaminoglycan_attachment_site
 FT modified_site 57..59
 FT /label= N-glycosylation_site
 FT region 282..291
 FT /label= basic_amino_acid_cluster
 FT domain 292..425
 FT /label= Cysteine-rich
 FT modified_site 360..362
 FT /label= N-glycosylation_site
 FT modified_site 372..374
 FT /label= N-glycosylation_site
 FT domain 426..449
 FT /label= transmembrane
 FT modified_site 465
 FT /label= phosphorylation_site
 FT /note= "Recognised by Protein Kinase C"
 FT modified_site 521
 FT /label= phosphorylation_site
 FT /note= "Recognised by Protein Kinase C"

J09194502-A.
 29-JUL-1997.
 PD 13-NOV-1996; 317060.
 PF 13-NOV-1995; JP-318651.
 PR (SEGK) SEIRAGAKU KOGYO CO LTD.
 PA WPI; 97-431470/40.
 DR N-PSDB; T86264.
 DR New chondroitin sulphate proteoglycan - consists of a 514 amino acid
 FT glyco:protein and 30 kD chondroitin sulphate which can be removed by
 PT neuraminidase or by O- or N-glycosidase
 PS Claim 1; Page 14-17; 23pp; Japanese.

CC A new chondroitin sulphate proteoglycan consists of the mature
 CC neuroglycan C (NGC) protein isolated from rat brain (and having
 CC the present sequence) and a sugar chain containing chondroitin
 CC sulphate of molecular weight ca. 30 kD and which can be isolated by
 CC reacting neuraminidase, O-glycosidase or N-glycosidase with NGC.
 CC The rat NGC protein also has a cysteine-rich domain, separated from
 CC the chondroitin sulphate domain by a cluster of basic amino acids.
 CC The chondroitin sulphate domain is itself separated from the cytoplasmic
 CC domain (containing two phosphorylation sites) by a transmembrane
 CC domain. The chondroitin sulphate proteoglycan can be used for
 CC developing new methods of diagnosis for diseases of the nervous
 CC system, as well as for their prevention and treatment.
 SQ Sequence 544 AA;

Query Match 65.5%; Score 36; DB 1; Length 544;
 Best Local Similarity 75.0%; Pred. No. 69;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AWTADGT 8
 I | | | | |
 Db 105 AWTADGT 112

RESULT 4
 R15625 standard; Protein; 782 AA.
 ID R15625;
 AC R15625;
 DT 17-MAR-1992 (first entry)
 DE Cellulase AB-1.
 KW Detergents; pharmaceuticals; deinking; carboxymethylcellulose.
 OS Aeromonas strain no. 212.
 PN J03251174-A.
 PD 08-NOV-1991.
 PF 28-FEB-1990; 045465.
 PR 28-FEB-1990; JP-045465.
 PA (OJIP) OJI PAPER KK.
 DR WPI; 91-373412/51.
 DR N-PSDB; Q15178.
 PT Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -
 PT of opt. pH when carboxymethylcellulose is used as substrate.
 PS Claim 2; Fig 3; 8pp; Japanese.
 CC The sequence was deduced from the gene which was sequenced from
 CC plasmid, pAEC 1, prep. by ligating chromosomal DNA contg. the
 CC gene (obtd. from Aeromonas) into pUC18. The protein has amol. wt.
 CC of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-
 CC methylcellulose is the substrate. The N-terminal sequence: GIHADT-
 CC has been confirmed by Edman degradation. The gene can be used to
 CC produce recombinant enzyme which is used for the effective utilis-
 CC ation of biomass resources and the mfr. of pharmaceuticals and
 CC foodstuffs, and also for the detergent and deinking of waste paper.
 SQ Sequence 782 AA;

Query Match 63.6%; Score 35; DB 1; Length 782;
 Best Local Similarity 50.0%; Pred. No. 1.6e+02;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AWTADGTN 10
 I | | | | |
 Db 574 AWTADGTN 583

RESULT 5
 R31372 standard; Protein; 428 AA.
 ID R31372;
 AC R31372;
 DT 24-JUN-1993 (first entry)
 DE Murine bone morphogenic protein-9 (BMP-9).
 KW bone formation; cartilage formation; wound healing; tissue repair;
 KW surgery; fracture treatment; periodontal disease; osteoporosis;
 KW increase neuronal survival; transplantation; nerves; nervous system.
 OS Mus musculus.

FH Key Location/Qualifiers
 FT 319. .427
 FT /label= mature peptide
 PN W09300432-A.
 PD 07-JAN-1993.
 PE 25-JUN-1992; U05374.
 PR (GENY) GENETICS INST INC.
 PA Celeste AJ, Wozney JM;
 PI WPI; 93-036379/04.
 DR N-PSDB; Q35243.
 PT New BMP-9 polypeptide(s) induce bone and cartilage formation -
 PT used for treating osteoporosis and fractures, healing wounds and
 PT increasing neuronal survival
 PS Claim 1; Fig 1; 60pp; English.
 CC This sequence is thought to be the primary translation product of
 CC murine bone morphogenic protein. Based on the knowledge of other
 CC BMP's and proteins within the TGF-Beta family it is predicted that
 CC the precursor polypeptide would be cleaved at the multibasic
 CC sequence Arg-Arg-Iys-Arg in agreement with the proposed consensus
 CC proteolytic processing sequence of Arg-X-X-Arg to generate a 110
 CC amino acid mature peptide. It is expected that processing into the
 CC mature form will involve dimerisation and removal of the N-terminal
 CC region in a manner analogous to the related protein TGF-beta,
 CC giving a homodimer of 2 subunits each with a molecular weight of
 CC approx. 12,000 Da. The protein can be used to induce bone and/or
 CC cartilage formation and in wound healing and tissue repair. It
 CC can be used in surgery of the treatment of fractures, periodontal
 CC disease or osteoporosis. The protein can also increase neuronal
 CC survival and can be used in transplantation and treatment of
 CC conditions exhibiting a decrease in neuronal survival.
 SQ Sequence 428 AA;

Query Match 63.6%; Score 35; DB 1; Length 428;
 Best Local Similarity 66.7%; Pred. No. 79;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
 II II II
 Db 213 WVRADSTTN 221
 RESULT 6
 ID R86905 standard; Protein: 428 AA.
 AC R86905;
 DT 10-MAY-1996 (first entry)
 DE Murine BMP-9.
 KW BMP-9; bone; cartilage; wound healing; liver; tissue repair.
 OS Mus musculus strain C57B46xCBA.
 FH Key Location/Qualifiers
 FT cleavage_site 318. .319
 FT /note= "proteolytic cleavage site"
 FT 319. .428
 FT /label= Mat_protein
 FT W09533830-A1.
 PN 14-DEC-1995.
 PE 05-JUN-1995; U07084.
 PR 06-JUN-1994; US-254353.
 PA (GENY) GENETICS INST INC.
 PI Celeste AJ, Rosen VA, Song JJ, Thies SR, Wozney JM;
 DR WPI; 96-040235/04.
 DR N-PSDB; Q74084.
 PT BMP-9 protein able to induce bone and cartilage formation - also
 PT useful in wound healing, tissue repair and for inducing hepatic
 PT growth and function
 PS Example 1; Page 36-38; 75pp; English.
 CC Murine BMP-9 protein (R86905) is the product of a cDNA clone.
 CC ML14a (Q74084), isolated from a mouse liver cDNA library. The
 CC primary translation product is cleaved to yield a 110-amino acid
 CC protein that forms a mature, homodimeric, active species.
 CC Recombinant BMP-9 is obtd. by expression of the cDNA in procaryotic

CC or eucaryotic hosts. It has therapeutic applns. in the
 CC formation of bone and cartilage, wound healing and tissue repair.
 SQ Sequence 428 AA;

Query Match 63.6%; Score 35; DB 1; Length 428;
 Best Local Similarity 66.7%; Pred. No. 79;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
 II II II
 Db 213 WVRADSTTN 221
 RESULT 7
 ID W02256 standard; Protein: 572 AA.
 AC W02256;
 DT 25-APR-1997 (first entry)
 DE Mouse L-glutamate transporter protein (pGTA1).
 KW Mouse; murine; cerebellum; L-glutamate; transporter; nerve; cell;
 KW development; disorder; tissue damage; cerebral infarction;
 KW gene therapy; treatment; Xenopus; egg cell; drug screening.
 OS Mus musculus.
 PN W09626957-A1.
 PD 06-SEP-1996.
 PR 23-FEB-1996; J00413.
 PR 27-FEB-1995; JP-037848.
 PA (HISM) HISAMITSU PHARM CO LTD.
 PI Tanaka K, Wada K;
 FI WPI; 96-412737/41.
 DR N-PSDB; T62507.
 PT Nerve tissue derived glutamate transporter protein - useful in study
 PT and diagnosis of processes occurring in nerve cell development
 PT disorders, and nerve tissue damage after cerebral infarction
 PS Claim 1; Pages 12-14; 28pp; Japanese.
 CC The present sequence is a mouse cerebellar tissue derived
 CC L-glutamate transporter protein. The protein and its cDNA can be
 CC used in the study of nerve cell development disorders and nerve
 CC tissue damage after cerebral infarction, etc., and in the
 CC production of gene therapies and drugs for the treatment of such
 CC disorders. Also Xenopus egg cells transformed with the cDNA can be
 CC used to screen glutamate transporter function affecting drugs.
 CC RNA isolated from ICR mouse cerebellar tissue was used to
 CC construct a cDNA library, which was screened using rat glutamate
 CC transporter GLT1 fragments. An active clone, pGTA1, was generated
 CC in pBluescript SK(-) and KpnI cleaved. The DNA fragment obtained
 CC was used to generate mRNA (MGLT1), which was injected directly
 CC into Xenopus egg cells. These were cultured in Barth medium, with
 CC labelled glutamate so that transport into the cells could be
 CC studied. In the presence of 115 mM Na+ or Li+, transport of
 CC L-glutamate into the cells was 670 and 37 pmol/cell/hour,
 CC respectively. When D-glutamate was used no transport occurred.
 SQ Sequence 572 AA;

Query Match 63.6%; Score 35; DB 1; Length 572;
 Best Local Similarity 60.0%; Pred. No. 1.le+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWVTADGTVN 10
 II II II
 Db 353 AWITALGTAS 362
 RESULT 8
 ID W26598 standard; Protein: 574 AA.
 AC W26598;
 DT 23-JAN-1998 (first entry)
 DE Human glutamate transporter protein EAAT2.
 KW EAAT2; excitatory amino acid transporter protein;
 KW glutamate transporter; human; glutamic acid; agonist; antagonist;

KW antibody.
 OS Homo sapiens.
 PN US558782-A.
 PD 19-AUG-1997.
 PF 20-OCT-1993; 140729.
 PR 20-OCT-1993; US-140729.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 PI Amara SG, Arriza JL;
 DR WPI; 97-424239/39.
 DR N-PSDB; T90412.
 PT Human glutamate transporter protein EAAT2 and related nucleic acids
 PT - useful for screening for (antagonist or analogues that modulate
 PT glutamate transport
 PS Claim 3; Column 37-42; 70pp; English.
 CC This polypeptide comprises the novel human excitatory amino acid
 CC (glutamate) transporter EAAT2. Its amino acid sequence was deduced
 CC from a cDNA clone (see T90412) obtained from a human motor cortex
 CC cDNA library. Human amino acid transporter proteins EAAT1
 CC (see W26600), EAAT3 (see W26601) and ASC1 (W26599) have also been
 CC identified. The amino acid transporters can be expressed in
 CC transformed prokaryotic or eukaryotic host cells. The ability to
 CC synthesize human transporter molecules provides an efficient means
 CC for rational drug design and rapid screening of potentially useful
 CC compounds that have transport-modulating properties. Synthetic
 CC peptides generated from the proteins can be used as antigens or
 CC as competitors/inhibitors of (antagonist binding to the
 CC transporter proteins. The proteins can also be used to raise
 CC antibodies.
 CC Sequence 574 AA;

Query Match 63.6%; Score 35; DB 1; Length 574;
 Best Local Similarity 60.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWTADGTGVN 10
 II:III II:
 Db 354 AWITALGTAS 363

RESULT 9
 W58553
 ID W58553 standard; Protein; 574 AA.
 AC W58553;
 DT 04-SEP-1998 (first entry)
 DE Human excitatory amino acid transporter EAAT2.
 DE Human excitatory amino acid transporter; EAAT1; EAAT2; EAAT3; ASC1;
 KW neutral amino acid transporter; diagnosis; genetic disease.
 OS Homo sapiens.
 PN US5776774-A.
 PD 07-JUL-1998.
 PF 23-OCT-1995; 546666.
 PR 20-OCT-1993; US-140729.
 PR 23-OCT-1995; US-546666.
 PA (OREG-) STATE OF OREGON.
 PI Amara SG, Arriza JL;
 DR WPI; 98-398047/34.
 DR N-PSDB; V31149.
 PT New nucleic acid encoding human excitatory amino acid transporter 3
 PT - and related vectors and transformed cells, useful for diagnosis of
 PT genetic disease, production of antibodies, drug screening etc.
 PS Example 2; Fig 3; 72pp; English.
 CC The present sequence represents human excitatory amino acid transporter 2
 CC (EAAT2) from an example of the present invention. The present invention
 CC also describes human excitatory amino acid transporter 3 (EAAT3). Probes
 CC derived from the nucleic acid encoding EAAT3 are used to detect pattern
 CC and extent of transporter gene expression, e.g. for diagnosis of genetic
 CC disease and detection of related gene sequences. Also natural and
 CC synthetic peptides derived from EAAT3 are used to raise antibodies (AB);
 CC as competitors of EAAT3 and to produce inhibitors of binding of
 CC (antagonists (or their analogues) to EAAT3. Cells transformed with a
 CC vector containing the nucleic acid encoding EAAT3 are used to produce
 CC recombinant EAAT3 and to screen for modulators of transporter activity.

SQ Sequence 574 AA;

Query Match 63.6%; Score 35; DB 1; Length 574;
 Best Local Similarity 60.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWTADGTGVN 10
 II:III II:
 Db 354 AWITALGTAS 363

RESULT 10
 W83923
 ID W83923 standard; Protein; 574 AA.
 AC W83923;
 DT 29-JAN-1999 (first entry)
 DE Human excitatory amino acid transporter EAAT2.
 DE Amino acid transporter; EAAT2; drug screening; assay; inhibitor;
 KW excitatory; recombinant; gene therapy; human.
 OS Homo sapiens.
 PN US5840516-A.
 PD 24-NOV-1998.
 PF 19-AUG-1997; 916745.
 PR 20-OCT-1993; US-140729.
 PR 19-AUG-1997; US-916745.
 PA (OREG-) STATE OF OREGON.
 PI Amara SG, Arriza JL;
 DR WPI; 99-034033/03.
 DR N-PSDB; V69231.
 PT Screening assay for inhibitors of excitatory amino acid transport -
 PT using cells expressing recombinant excitatory amino acid transporter
 PT EAAT2
 PS Example 2; Fig 3A-E; 72pp; English.
 CC This represents a human excitatory amino acid transporter EAAT2. The
 CC invention relates to novel amino acid transporters (W83921 to W83924)
 CC and genes (V69229 to V69232) encoding the transporters respectively.
 CC The invention provides a screening assay for inhibitors of excitatory
 CC amino acid transporter that comprises transforming a cell culture with
 CC a recombinant expression construct that expresses the human excitatory
 CC amino acid transporter EAAT2, and assaying the cell culture with a test
 CC compound to determine if (or the extent to which) the compound inhibits
 CC EAAT2-mediated excitatory amino acid transport. The recombinant
 CC expression constructs are useful for transforming cells, which do not
 CC ordinarily express a particular amino acid transporter, into cell lines
 CC expressing the receptor. These cells are useful as intermediates for
 CC making cell membrane preparations which are useful for transporter
 CC activity assays and in drug screening. The recombinant expression
 CC constructs may also be useful in gene therapy or for site-directed
 CC mutagenesis.
 CC Sequence 574 AA;

Query Match 63.6%; Score 35; DB 1; Length 574;
 Best Local Similarity 60.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWTADGTGVN 10
 II:III II:
 Db 354 AWITALGTAS 363

RESULT 11
 Y13394
 ID Y13394 standard; Protein; 640 AA.
 AC Y13394;
 DT 25-JUN-1999 (first entry)
 DE Amino acid sequence of protein PRO331.
 DE Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW congenital microvillus atrophy; skin disease; cell growth;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;

KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
 KW anti-thrombotic; wound healing; tissue repair.
 OS Homo sapiens.
 PN WO9914328-A2.
 PD 25-MAR-1999. U19330.
 PF 16-SEP-1998; US-066840.
 PR 25-NOV-1997; US-059113.
 PR 17-SEP-1997; US-059117.
 PR 17-SEP-1997; US-059115.
 PR 17-SEP-1997; US-059119.
 PR 17-SEP-1997; US-059121.
 PR 17-SEP-1997; US-059122.
 PR 17-SEP-1997; US-059184.
 PR 18-SEP-1997; US-059263.
 PR 18-SEP-1997; US-059266.
 PR 15-OCT-1997; US-062125.
 PR 17-OCT-1997; US-062285.
 PR 17-OCT-1997; US-062287.
 PR 21-OCT-1997; US-063486.
 PR 24-OCT-1997; US-062814.
 PR 24-OCT-1997; US-062816.
 PR 24-OCT-1997; US-063045.
 PR 24-OCT-1997; US-063120.
 PR 24-OCT-1997; US-063121.
 PR 24-OCT-1997; US-063127.
 PR 24-OCT-1997; US-063128.
 PR 27-OCT-1997; US-063329.
 PR 27-OCT-1997; US-063327.
 PR 28-OCT-1997; US-063541.
 PR 28-OCT-1997; US-063542.
 PR 28-OCT-1997; US-063544.
 PR 28-OCT-1997; US-063549.
 PR 28-OCT-1997; US-063550.
 PR 28-OCT-1997; US-063564.
 PR 28-OCT-1997; US-063565.
 PR 29-OCT-1997; US-063435.
 PR 29-OCT-1997; US-063704.
 PR 29-OCT-1997; US-063732.
 PR 29-OCT-1997; US-063738.
 PR 29-OCT-1997; US-063734.
 PR 29-OCT-1997; US-064215.
 PR 29-OCT-1997; US-064215.
 PR 31-OCT-1997; US-063870.
 PR 31-OCT-1997; US-064103.
 PR 03-NOV-1997; US-064248.
 PR 07-NOV-1997; US-064809.
 PR 12-NOV-1997; US-065186.
 PR 17-NOV-1997; US-065846.
 PR 18-NOV-1997; US-065693.
 PR 21-NOV-1997; US-066120.
 PR 21-NOV-1997; US-066364.
 PR 24-NOV-1997; US-066772.
 PR 24-NOV-1997; US-066466.
 PR 24-NOV-1997; US-066770.
 PR 24-NOV-1997; US-066511.
 PR 24-NOV-1997; US-066453.
 PR (GEN J) GENENTECH INC.
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WL, Yuan J;
 DR WPI: 99-229533/19.
 DR N-PSDB: X52265.
 DR New isolated human genes and polypeptides used in, e.g. treatment of
 PT gastrointestinal ulceration
 PS Claim 12; Fig 104; 320pp; English.
 PS Y13344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial

CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to growth
 CC or survival of nerve cells including parkinson's disease, Alzheimer's
 CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
 CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair; PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 SQ Sequence 640 AA;

Query Match 63.6%; Score 35; DB 1; Length 640;
 Best Local Similarity 55.6%; Pred No. 1.2e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 AWVTADGTV 9

:|:| :|:|

Db 385 SWTPNGTV 393

RESULT 12

R98811

ID R98811 standard; Protein: 2273 AA.

AC R98811;

DT 10-FEB-1997 (first entry)

DE Erysiphe graminis acetyl coenzyme A carboxylase.

KW Acetyl coenzyme A carboxylase; ACoAcase; powdery mildew; fungus;

OS ACCase; fatty acid biosynthesis; EC 6.4.1.2.

PN FR2727129-A1.

PD 24-MAY-1996.

PF 21-NOV-1994; 014187.

PR 21-NOV-1994; FR-014187.

PA (RHON) RHONE POULENC AGROCHIMIE.

PI Grosjean CMC, Hollomon DW, Lebrun M;

DR WPI: 96-270416/28.

DR N-PSDB: T30758.

PT Microorganism with specific biochemical activity deleted by mutation

PT - and complemented, used in system to identify cpds with plant

PT protecting activity, also new gene for acetyl coenzyme A carboxylase

PS Example 2; Page 16-22; 26pp; French.

CC The present sequence is the deduced amino acid sequence of acetyl

CC coenzyme A carboxylase (ACoAcase, EC 6.4.1.2) from powdery mildew

CC DNA sequence isolated from an EcoRI genomic library by screening with

CC a fragment of the Saccharomyces cerevisiae ACoAcase gene. The amino

CC acid sequence has 63 % identity and 77 % similarity to the yeast

CC enzyme; it also has 47 % identity (67 % similarity) with the

CC ACoAcase proteins from rat and chicken. ACoAcase activity is the

CC preferred target biochemical activity in a new screening system.

SQ Sequence 2273 AA;

Query Match 61.8%; Score 34; DB 1; Length 2273;
 Best Local Similarity 50.0%; Pred. No. 8e+02;

Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 AWVTADGTVN 10

:|:| :|:|

Db 2035 SNWVDPTIN 2044

RESULT 13

W08333

ID W08333 standard; Protein: 2089 AA.

AC W08333;

DT 12-FEB-1997 (first entry)

DE Cyclotella cryptica acetyl-coenzyme A carboxylase.

KW Gene therapy; herbicide resistance; fatty acid biosynthesis;

KW transgene; recombinant; sunflower; safflower; rape; Olive; soybean;

KW peanut; flax; castor; palm; coconut; cotton; oil.
 OS Cyclostella cryptica T31L.
 FH Key Location/Qualifiers
 FT binding_site 287..307
 FT /note= "putative ANP binding site"
 FT binding_site 1476..1526
 FT /note= "putative carboxy-biotin binding site"
 FT binding_site 1758..1777
 FT /note= "putative acetyl-CoA binding site"
 PN US5559220-A.
 PD 24-SEP-1996.
 PF 14-SEP-1993; 120938.
 PR 14-SEP-1993; US-120938.
 PR 07-APR-1995; US-418893.
 PA (MIDE) MIDWEST RES INST.
 PI Ohlrogge JB, Roessler PG;
 DR WPI; 96-442460/44.
 DR N-PSDB; T45727, T45728.
 PT Acetyl-coenzyme A carboxylase gene from Cyclostella cryptica - is
 used to impart herbicide resistance and for increasing fatty acid
 content of organisms transformed with the DNA
 PS Claim 2; Column 41-52; 40pp; English.
 CC W08333 is the acetyl-coenzyme A carboxylase (ACAC) enzyme from the
 CC photosynthetic diatom, Cyclostella cryptica. ACAC catalyses the first
 CC regulatory or rate-limiting step in fatty acid biosynthesis in bacteria,
 CC animals, plants and yeast. The enzyme catalyses the carboxylation of
 CC acetyl CoA to form malonyl CoA. Unlike ACAC from monocotyledonous
 CC plants, C. cryptica ACAC is not strongly inhibited by herbicides,
 CC because of this fact the ACAC gene may be used in gene therapy for the
 CC production of plants resistant to certain herbicides. Such plants
 CC may be cultivated using stronger herbicides that would normally damage
 CC or kill the plants. Also the C. cryptica ACAC gene may be inserted into
 CC a plant's genome to increase their fatty acid content, to produce
 CC large quantities of lipids, esp. triglycerides, at low cost. Major
 CC crops grown primarily or secondarily for their lipids include rape,
 CC sunflower, safflower, olive, soybean, castor, peanut, palm, coconut,
 CC flax and cotton.
 SQ Sequence 2089 AA;

Query Match 61.8%; Score 34; DB 1; Length 2089;
 Best Local Similarity 50.0%; Pred. No. 7.3e+02;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 AWTADGTN 10
 :||| | | | |
 DB 1882 SWVVDPITN 1891

RESULT 14
 W99396
 ID W99396 standard; Protein; 401 AA.
 AC W99396;
 DT 08-JUN-1999 (first entry)
 DE S. antibioticus olepl gene product.
 KW Gene cluster; bacterium; enzyme; macrolide; antibiotic; oleandomycin;
 KW secondary metabolite; hybridisation; probe; glycosylation; macrolactone.
 OS Streptomyces antibioticus.
 PN W09905283-A2.
 PD 04-FEB-1999.
 PF 21-JUL-1998; F01593.
 PR 12-JUN-1998; FR-007411.
 PR 25-JUL-1997; FR-009458.
 PA (HMRI) HOECHST MARION ROUSSEL.
 PI Cortes J, Gaiser S, Leadlay P, Michel JM, Raynal MC,
 PI Salah-Bey K, Fromentin C, Mendez C, Salas JA;
 DR WPI; 99-142938/12.
 DR N-PSDB; X25775.
 PT New nucleic acid sequences encoding enzymes involved in macrolide
 PT biosynthesis - useful for producing hybrid secondary metabolites,
 PT particularly erythromycin analogues
 PS Disclosure; Fig 22; 221pp; French.
 CC This sequence represents the product encoded by the olepl gene from

the olepl-olev gene cluster from the bacterium Streptomyces antibioticus.
 The ole gene cluster encodes enzymes involved in the production of the
 macrolide antibiotic oleandomycin as a secondary metabolite. The genes
 are used to produce hybrid secondary metabolites in S. antibioticus,
 i.e. oleandomycin analogues which may have improved properties or as
 hybridisation probes for isolating homologous genes involved in
 glycosylation of macrolactones in macrolide-producing strains.
 SQ Sequence 401 AA;

Query Match 61.8%; Score 34; DB 1; Length 401;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AWTAD 6
 :||| | | | |
 DB 58 AWTAD 63

RESULT 15
 Y01648
 ID Y01648 standard; Protein; 541 AA.
 AC Y01648;
 DT 23-JUN-1999 (first entry)
 DE Cytochrome P450 encoded by a codon modified cDNA.
 KW Cytochrome P450; protein expression; yeast; mutant.
 OS Synthetic.
 OS Trifolium aestivum.
 PN FR2768748-Al.
 PD 26-MAR-1999.
 PF 24-SEP-1997; 012094.
 PR 24-SEP-1997; FR-012094.
 PA (RHON) RHONE-POULENC AGROCHIMIE.
 PI Batard Y, Durst F, Schalk M, Werck RD;
 DR WPI; 99-217499/19.
 DR N-PSDB; X26878.
 PT DNA for expression in yeasts - with codon changes based on yeast
 PT codon usage
 PS Example 4; Page 24-26; 31pp; French.
 CC The present sequence represents a cytochrome P450, encoded for by a
 CC cDNA which was created by altering the codons of X26865 to exemplify
 CC the invention. The specification describes a DNA sequence that encodes
 CC a protein of interest and contains regions with a high content of codons
 CC poorly suited to yeasts, where a sufficient number of these codons
 CC are replaced by corresponding codons suited to yeasts in these
 CC regions. Yeasts transformed with vectors such DNA sequences
 CC can be cultured to produce the protein of interest (especially a
 CC plant protein) or, when the protein is an enzyme (such as cytochrome
 CC P450), can be cultured in the presence of a substrate for the enzyme
 CC to produce a substrate conversion product.
 SQ Sequence 541 AA;

Query Match 61.8%; Score 34; DB 1; Length 541;
 Best Local Similarity 71.4%; Pred. No. 1.6e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 WYTADGT 8
 :|::| | | | |
 DB 434 WLSADGT 440

Search completed: November 13, 1999, 18:59:54
 JOB time: 133 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:30 ; Search time 64.87 Seconds
(without alignments)
1.762 Million cell updates/sec

Title: US-08-913-430-8
Perfect score: 55
Sequence: 1 ANVTADGTGN 10

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/PCTUS9_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	63.6	574	1 US-08-140-729A-7	Sequence 7, Appl
2	35	63.6	428	1 US-08-050-132A-2	Sequence 2, Appl
3	35	63.6	574	2 US-08-546-666-7	Sequence 7, Appl
4	35	63.6	574	2 US-08-916-745-7	Sequence 7, Appl
5	35	63.6	574	2 US-08-948-569A-4	Sequence 4, Appl
6	35	63.6	574	2 US-08-663-808-6	Sequence 6, Appl
7	35	63.6	574	2 US-09-042-929-7	Sequence 7, Appl
8	35	63.6	574	2 US-08-546-661-7	Sequence 7, Appl
9	35	63.6	574	2 US-09-042-960-7	Sequence 7, Appl
10	35	63.6	428	3 PCT-US92-05374A-2	Sequence 2, Appl
11	35	63.6	428	3 PCT-US95-07084-2	Sequence 2, Appl
12	34	61.8	2089	1 US-08-418-893D-23	Sequence 23, Appl
13	34	61.8	2089	1 US-08-418-893D-24	Sequence 24, Appl
14	34	61.8	2237	1 US-08-354-973-1	Sequence 1, Appl
15	34	61.8	1250	2 US-08-441-139-9	Sequence 9, Appl
16	33	60.0	377	1 US-08-480-882B-6	Sequence 6, Appl
17	33	60.0	377	2 US-08-480-210-6	Sequence 6, Appl
18	33	60.0	377	2 US-08-220-401-4	Sequence 4, Appl
19	33	60.0	2257	2 US-08-611-107-10	Sequence 10, Appl
20	33	60.0	2172	2 US-08-611-107-31	Sequence 31, Appl
21	33	60.0	377	2 US-08-437-362-4	Sequence 4, Appl
22	33	60.0	1313	2 US-08-244-537-2	Sequence 2, Appl
23	33	60.0	2257	2 US-08-422-560A-10	Sequence 10, Appl
24	33	60.0	2254	2 US-08-677-010-3	Sequence 3, Appl
25	33	60.0	429	2 US-09-074-512-1	Sequence 1, Appl
26	32	58.2	221	1 US-07-949-812-2	Sequence 2, Appl
27	32	58.2	192	1 US-07-949-812-3	Sequence 3, Appl
28	32	58.2	2199	3 PCT-US95-11684-2	Sequence 2, Appl
29	31	56.4	883	1 US-08-106-433A-2	Sequence 2, Appl
30	31	56.4	2710	1 US-08-480-604A-6	Sequence 6, Appl
31	31	56.4	811	1 US-08-480-604A-7	Sequence 7, Appl
32	31	56.4	812	1 US-08-480-604A-29	Sequence 29, Appl
33	31	56.4	596	1 US-08-565-386-11	Sequence 11, Appl
34	31	56.4	330	2 US-08-671-268A-18	Sequence 18, Appl
35	31	56.4	2710	2 US-08-405-496A-6	Sequence 6, Appl
36	31	56.4	811	2 US-08-405-496A-7	Sequence 7, Appl
37	31	56.4	435	2 US-08-338-530A-2	Sequence 2, Appl
38	31	56.4	166	3 PCT-US93-02869-8	Sequence 8, Appl
39	31	56.4	1833	3 PCT-US95-02251-18	Sequence 18, Appl

Sequence 2, Appl
Sequence 3, Appl
Sequence 35, Appl
Sequence 5, Appl
Sequence 20, Appl
Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-08-140-729A-7
; Sequence 7, Application US/08140729A
; Patent No. 5658782
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Ariza, Jeffrey L
; TITLE OF INVENTION: Amino Acid Transporters and Uses
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/140,729A
; FILING DATE: 20 OCT 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5658782nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-140-729A-7

Query Match 63.6%; Score 35; DB 1; Length 574;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ANVTADGTGN 10
   |||||
DB 354 AWITALGTAS 363

RESULT 2
US-08-050-132A-2
; Sequence 2, Application US/08050132A
; Patent No. 5661007
; GENERAL INFORMATION:
; APPLICANT: Wozney, John M.
; APPLICANT: Celeste, Anthony
; TITLE OF INVENTION: BMP-9 COMPOSITIONS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
```


STREET: Legal Affairs - 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: US
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/050,132A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kapinos, Ellen J.
REGISTRATION NUMBER: 32,245
REFERENCE/DOCKET NUMBER: GI 5186A
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 428 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-050-132A-2

Query Match 63.6%; Score 35; DB 1; Length 428;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Caps 0;

QY 2 WVTADGTVN 10
||| ||| |
Db 213 WVRADSTN 221

RESULT 3
US-08-546-666-7
Sequence 7, Application US/08546666
Patent No. 5776774
GENERAL INFORMATION:
APPLICANT: Amara, Susan G
TITLE OF INVENTION: Amino Acid Transporters and Uses
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/546,666
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/140,729
FILING DATE: 20 OCT 1993
ATTORNEY/AGENT INFORMATION:
NAME: No. 5776774nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 93,509
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 574 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-546-666-7

Query Match 63.6%; Score 35; DB 2; Length 574;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWVTADGTVN 10
||| ||| |
Db 354 AWITAGTAS 363

RESULT 4
US-08-916-745-7
Sequence 7, Application US/08916745
Patent No. 5840516
GENERAL INFORMATION:
APPLICANT: Amara, Susan G
TITLE OF INVENTION: Amino Acid Transporters and Uses
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,745
FILING DATE: 19-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/140,729
FILING DATE: 20-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: No. 5840516nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 93,509
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 574 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-916-745-7

Query Match 63.6%; Score 35; DB 2; Length 574;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWVTADGTVN 10
||| ||| |
Db 354 AWITAGTAS 363

RESULT 5
US-08-948-569A-4
; Sequence 4, Application US/08948569A
; Patent No. 5882926
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; APPLICANT: Eliasof, Scott
; APPLICANT: Kavanaugh, Michael P
; TITLE OF INVENTION: Excitatory Amino Acid Transporter Genes
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,569A
; FILING DATE: 10-OCT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5882926nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509-F
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-948-569A-4

Query Match 63.6%; Score 35; DB 2; Length 574;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWTADGTVN 10
||:|||||
Db 354 AWITALGTAS 363

RESULT 6
US-08-663-808-6
; Sequence 6, Application US/08663808
; Patent No. 5912171
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; APPLICANT: Fairman, Wendy A
; TITLE OF INVENTION: Amino Acid Transporters and Uses
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,808
; FILING DATE:

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5912171nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-663-808-6

Query Match 63.6%; Score 35; DB 2; Length 574;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWTADGTVN 10
||:|||||
Db 354 AWITALGTAS 363

RESULT 7
US-09-042-929-7
; Sequence 7, Application US/09042929
; Patent No. 5919628
; GENERAL INFORMATION:
; APPLICANT: Amara, Susan G
; APPLICANT: Arriza, Jeffrey L
; TITLE OF INVENTION: Amino Acid Transporters and Uses
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/042,929
; FILING DATE:

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/140,729
; FILING DATE: 20 OCT 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5919628nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,509
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 574 amino acids
; TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-042-929-7

Query Match 63.6%; Score 35; DB 2; Length 574;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWTADGTVN 10
||:|||||
Db 354 AWITALGTAS 363

RESULT 8

US-08-546-661-7
Sequence 7, Application US/08546661
Patent No. 5919699

GENERAL INFORMATION:
APPLICANT: Amara, Susan G

TITLE OF INVENTION: Amino Acid Transporters and Uses

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.

STREET: 10 South Wacker Drive, Suite 3000

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/546,661

FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/140,729

FILING DATE: 20 OCT 1993

ATTORNEY/AGENT INFORMATION:
NAME: No. 5919699nan, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 93,509

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 574 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-546-661-7

Query Match 63.6%; Score 35; DB 2; Length 574;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWTADGTVN 10
||:|||||
Db 354 AWITALGTAS 363

RESULT 9

US-09-042-960-7

Sequence 7, Application US/09042960

Patent No. 5932424

GENERAL INFORMATION:
APPLICANT: Amara, Susan G

TITLE OF INVENTION: Amino Acid Transporters and Uses

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.

STREET: 10 South Wacker Drive, Suite 3000

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/05374A

FILING DATE: 19920625

CLASSIFICATION:

APPLICANT: Amara, Susan G
APPLICANT: Arriza, Jeffrey L
TITLE OF INVENTION: Amino Acid Transporters and Uses

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.

STREET: 10 South Wacker Drive, Suite 3000

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,960

FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/140,729

FILING DATE: 20 OCT 1993

ATTORNEY/AGENT INFORMATION:
NAME: No. 5932424nan, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 93,509

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 574 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-042-960-7

Query Match 63.6%; Score 35; DB 2; Length 574;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWTADGTVN 10
||:|||||
Db 354 AWITALGTAS 363

RESULT 10

PCT-US92-05374A-2

Sequence 2, Application PC/TUS9205374A

GENERAL INFORMATION:
APPLICANT: Wozney, John M.

APPLICANT: Celeste, Anthony

TITLE OF INVENTION: BMP-9 COMPOSITIONS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.

STREET: Legal Affairs - 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: US

ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/05374A

FILING DATE: 19920625

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Kapinos, Ellen J.
REGISTRATION NUMBER: 32,245
REFERENCE/DOCKET NUMBER: GI 5186A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
LENGTH: 428 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-05374A-2

Query Match 63.6%; Score 35; DB 3; Length 428;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
|| || || ||
Db 213 WVRADSTTN 221

RESULT 11
PCT-US95-07084-2
Sequence 2, Application PC/TUS9507084
GENERAL INFORMATION:
APPLICANT: Rosen, Vicki A.
APPLICANT: Wozney, John M.
APPLICANT: Celeste, Anthony J.
TITLE OF INVENTION: BMP-9 COMPOSITIONS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: Legal Affairs - 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: US
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07084
FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kapinos, Ellen J.
REGISTRATION NUMBER: 32,245
REFERENCE/DOCKET NUMBER: GI 5186C-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1210
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
LENGTH: 428 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-07084-2

Query Match 63.6%; Score 35; DB 3; Length 428;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
|| || || ||
Db 213 WVRADSTTN 221

RESULT 12
US-08-418-893D-23
Sequence 23, Application US/08418893D
Patent No. 5559220
GENERAL INFORMATION:
APPLICANT: ROESSLER, PAUL G
APPLICANT: OHLROGGE, JOHN B
TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
STREET: 1617 Cole Blvd.
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,893D
FILING DATE: April 7, 1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/104,938
FILING DATE: September 14, 1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: O'CONNOR, EDNA
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON

TELECOMMUNICATION INFORMATION:
TELEPHONE: 303-231-1000
TELEFAX: 303-231-1098
TELEX:
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 2089 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-418-893D-23

Query Match 61.8%; Score 34; DB 1; Length 2089;
Best Local Similarity 50.0%; Pred. No. 5.2e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 AWVTADGTVN 10
|| || || ||
Db 1882 SWVVDPTN 1891

RESULT 13
US-08-418-893D-24
Sequence 24, Application US/08418893D
Patent No. 5559220
GENERAL INFORMATION:
APPLICANT: ROESSLER, PAUL G
APPLICANT: OHLROGGE, JOHN B
TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY

STREET: 1617 Cole Blvd.
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,893D
FILING DATE: April 7, 1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/104,938
FILING DATE: September 14, 1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: O'CONNOR, EDNA
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303-231-1000
TELEFAX: 303-231-1098
TELEX:

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:
LENGTH: 2089 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-418-893D-24

Query Match 61.8%; Score 34; DB 1; Length 2089;
Best Local Similarity 50.0%; Pred. No. 5.2e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AWTADGTVN 10
Db 1882 SWVVDPTN 1891

RESULT 14
US-08-354-973-1
Sequence 1, Application US/08354973
Patent No. 5641666
GENERAL INFORMATION:
APPLICANT: Vahlensieck, Hans-Friedrich
APPLICANT: Hinnen, Albert
TITLE OF INVENTION: Fungi Resistant to Soraphen A
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: Patent Dept., 520 White Plains Rd., POB 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591-9005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/354,973
FILING DATE: 13-DEC-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: PE/5-19802/A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2237 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
US-08-354-973-1

Query Match 61.8%; Score 34; DB 1; Length 2237;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AWTADGTVN 10
Db 2003 SWVVDPTN 2012

RESULT 15
US-08-441-139-9
Sequence 9, Application US/08441139
Patent No. 5773245
GENERAL INFORMATION:
APPLICANT: Witttrup, Dr. Karl D.
APPLICANT: Robinson, Anne S.
TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/441,139
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/089,997
FILING DATE: 06-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8646
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1250 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-441-139-9

Query Match 61.8%; Score 34; DB 2; Length 1250;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 AWTADGTVN 10
|||:|:|:
Db 374 AWTADGALD 383

Search completed: November 13, 1999, 10:56:31
Job time: 1360 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:36 ; Search time 251.81 Seconds
(without alignments)
2.514 Million cell updates/sec

Title: US-08-913-430-8
Perfect score: 55
Sequence: 1 AWTADGTGN 10

Scoring table: BLOSUM62

Searched: 497124 seqs, 53304441 residues

Database : Pending_Patents_AA.*

- 1: /cgn2.6/ptodata/1/paa/PCUS9_COMB.pep.*
- 2: /cgn2.6/ptodata/1/paa/US07_COMB.pep.*
- 3: /cgn2.6/ptodata/1/paa/US080_COMB.pep.*
- 4: /cgn2.6/ptodata/1/paa/US081_COMB.pep.*
- 5: /cgn2.6/ptodata/1/paa/US082_COMB.pep.*
- 6: /cgn2.6/ptodata/1/paa/US083_COMB.pep.*
- 7: /cgn2.6/ptodata/1/paa/US084_COMB.pep.*
- 8: /cgn2.6/ptodata/1/paa/US085_COMB.pep.*
- 9: /cgn2.6/ptodata/1/paa/US086_COMB.pep.*
- 10: /cgn2.6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2.6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2.6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2.6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2.6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2.6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2.6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2.6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2.6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2.6/ptodata/1/paa/US094_COMB.pep.*
- 20: /cgn2.6/ptodata/1/paa/PCUS9_COMB.pep.*
- 21: /cgn2.6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 22: /cgn2.6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 23: /cgn2.6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 24: /cgn2.6/ptodata/1/paa/US06_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	100.0	10	13	US-08-913-430-8
2	41	74.5	21	13	US-08-913-430-9
3	40	72.7	109	23	US-09-417-507-24399
4	38.5	70.0	556	12	US-08-869-696-2
5	38.5	70.0	395	12	US-08-869-696-3
6	38.5	70.0	348	12	US-08-869-696-4
7	38.5	70.0	497	12	US-08-869-696-9
8	37	67.3	410	1	PCT-US98-25729-2
9	37	67.3	494	1	PCT-US98-25729-12
10	37	67.3	410	1	PCT-US98-25729-14
11	37	67.3	410	16	US-09-203-283-2
12	37	67.3	494	16	US-09-203-283-12
13	37	67.3	410	16	US-09-203-283-14
14	37	67.3	846	17	US-09-328-352-6648
15	36	65.5	227	19	US-60-128-476-4008
16	35	63.6	574	1	PCT-US99-05930-2
17	35	63.6	53	1	PCT-US99-05930-4
18	35	63.6	428	3	US-07-720-590-2
19	35	63.6	428	3	US-08-050-132-2

20	35	63.6	257	3	US-08-065-844A-9	Sequence 9, Appli
21	35	63.6	428	5	US-08-254-353-2	Sequence 2, Appli
22	35	63.6	428	5	US-08-254-353A-2	Sequence 2, Appli
23	35	63.6	428	11	US-08-750-222-2	Sequence 2, Appli
24	35	63.6	428	11	US-08-750-222A-2	Sequence 2, Appli
25	35	63.6	428	12	US-08-815-652-2	Sequence 2, Appli
26	35	63.6	428	12	US-08-815-652B-2	Sequence 2, Appli
27	35	63.6	640	13	US-08-955-557-12	Sequence 12, Appli
28	35	63.6	574	14	US-09-040-736-2	Sequence 4, Appli
29	35	63.6	53	14	US-09-040-736-4	Sequence 4, Appli
30	35	63.6	574	14	US-09-042-709-7	Sequence 7, Appli
31	35	63.6	574	14	US-09-042-913-7	Sequence 7, Appli
32	35	63.6	574	14	US-09-042-937-7	Sequence 7, Appli
33	35	63.6	574	15	US-09-188-469-4	Sequence 4, Appli
34	35	63.6	574	15	US-09-188-496-6	Sequence 6, Appli
35	35	63.6	574	15	US-09-198-650-7	Sequence 7, Appli
36	35	63.6	574	16	US-09-227-614-7	Sequence 7, Appli
37	35	63.6	574	17	US-09-332-740-6	Sequence 6, Appli
38	35	63.6	640	23	US-09-413-232-117	Sequence 117, App
39	35	63.6	1886	23	US-09-417-507-39902	Sequence 39902, A
40	34	61.8	1250	2	US-07-956-699-9	Sequence 9, Appli
41	34	61.8	1250	3	US-08-089-997-9	Sequence 9, Appli
42	34	61.8	2089	4	US-08-120-938A-23	Sequence 23, Appli
43	34	61.8	2089	4	US-08-120-938A-24	Sequence 24, Appli
44	34	61.8	483	14	US-09-001-403-26	Sequence 26, Appli
45	34	61.8	610	16	US-09-248-796-17030	Sequence 17030, A

ALIGNMENTS

RESULT 1
US-08-913-430-8
; Sequence 8, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0 - beta
; SEQ ID NO 8
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-8

Query Match 100.0%; Score 55; DB 13; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.00076;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AWTADGTGN 10
|||||
Db 1 AWTADGTGN 10

RESULT 2
US-08-913-430-9
; Sequence 9, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0


```
; CURRENT APPLICATION NUMBER: US/08/913.430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 9
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-9

Query Match      74.5%; Score 41; DB 13; Length 21;
Best Local Similarity 90.0%; Pred. No. 0.57;
Matches 9; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 1 AWTADGTGN 10
   ||| |||||
Db 1 AWTADGTGN 10

RESULT 3
US-09-417-507-24399
; Sequence 24399, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 24399
; LENGTH: 109
; TYPE: PRT
; ORGANISM: A.fumigatus
US-09-417-507-24399

Query Match      72.7%; Score 40; DB 23; Length 109;
Best Local Similarity 70.0%; Pred. No. 4.9;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWTADGTGN 10
   ||| |||||
Db 68 AWTADGELN 77

RESULT 4
US-08-969-696-2
; Sequence 2, Application US/08869696C
; GENERAL INFORMATION:
; APPLICANT: Cameron-Mills, Verena
; APPLICANT: Lok, Finn
; APPLICANT: Sinjorgo, Catharina Maria Cornelia
; APPLICANT: Van Den Dool, Ronald Tako Marinus
; APPLICANT: Caspers, Martinus Petrus Maria
; APPLICANT: Van Zeijl-Van Der Valk, Maria Joanna
; TITLE OF INVENTION: ARABINOXYLAN DEGRADATION
; FILE REFERENCE: 11225.01US01
; CURRENT APPLICATION NUMBER: US/08/869,696C
; CURRENT FILING DATE: 1997-06-05
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 556
; TYPE: PRT
; ORGANISM: barley
US-08-869-696-2
```

```
Query Match      70.0%; Score 38.5; DB 12; Length 556;
Best Local Similarity 81.8%; Pred. No. 52;
Matches 9; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AW-VTADGTGN 10
   ||| |||||
Db 477 AWLVADGTGN 487

RESULT 5
US-08-869-696-3
; Sequence 3, Application US/08869696C
; GENERAL INFORMATION:
; APPLICANT: Cameron-Mills, Verena
; APPLICANT: Lok, Finn
; APPLICANT: Sinjorgo, Catharina Maria Cornelia
; APPLICANT: Van Den Dool, Ronald Tako Marinus
; APPLICANT: Caspers, Martinus Petrus Maria
; APPLICANT: Van Zeijl-Van Der Valk, Maria Joanna
; TITLE OF INVENTION: ARABINOXYLAN DEGRADATION
; FILE REFERENCE: 11225.01US01
; CURRENT APPLICATION NUMBER: US/08/869,696C
; CURRENT FILING DATE: 1997-06-05
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 395
; TYPE: PRT
; ORGANISM: barley
US-08-869-696-3

Query Match      70.0%; Score 38.5; DB 12; Length 395;
Best Local Similarity 81.8%; Pred. No. 36;
Matches 9; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AW-VTADGTGN 10
   ||| |||||
Db 316 AWLVADGTGN 326

RESULT 6
US-08-869-696-4
; Sequence 4, Application US/08869696C
; GENERAL INFORMATION:
; APPLICANT: Cameron-Mills, Verena
; APPLICANT: Lok, Finn
; APPLICANT: Sinjorgo, Catharina Maria Cornelia
; APPLICANT: Van Den Dool, Ronald Tako Marinus
; APPLICANT: Caspers, Martinus Petrus Maria
; APPLICANT: Van Zeijl-Van Der Valk, Maria Joanna
; TITLE OF INVENTION: ARABINOXYLAN DEGRADATION
; FILE REFERENCE: 11225.01US01
; CURRENT APPLICATION NUMBER: US/08/869,696C
; CURRENT FILING DATE: 1997-06-05
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 348
; TYPE: PRT
; ORGANISM: barley
US-08-869-696-4

Query Match      70.0%; Score 38.5; DB 12; Length 348;
Best Local Similarity 81.8%; Pred. No. 32;
Matches 9; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AW-VTADGTGN 10
   ||| |||||
Db 269 AWLVADGTGN 279
```

```

RESULT 7
US-08-869-696-9
; Sequence 9, Application US/08869696C
; GENERAL INFORMATION:
; APPLICANT: Cameron-Mills, Verena
; APPLICANT: Lok, Finn
; APPLICANT: Sinjorgo, Catharina Maria Cornelia
; APPLICANT: Van Den Dool, Ronald Tako Marinus
; APPLICANT: Caspers, Martinus Petrus Maria
; APPLICANT: Van Zeijl-Van Der Valk, Maria Joanna
; TITLE OF INVENTION: ARABINOXYLAN DEGRADATION
; FILE REFERENCE: 11225.01US01
; CURRENT APPLICATION NUMBER: US/08/869,696C
; CURRENT FILING DATE: 1997-06-05
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 497
; TYPE: PRT
; ORGANISM: barley
US-08-869-696-9

Query Match 70.0%; Score 38.5; DB 12; Length 497;
Best Local Similarity 81.8%; Pred. No. 46;
Matches 9; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AW-VTADGTVN 10
Db 418 AMLVDADGTVN 428

RESULT 8
PCT-US98-25729-2
; Sequence 2, Application PC/TUS9825729
; GENERAL INFORMATION:
; APPLICANT: Wright C, Anita
; APPLICANT: Powell L, Jan
; APPLICANT: Morris J, Glenn
; APPLICANT: University of Maryland Biotechnology Institute
; TITLE OF INVENTION: V.Vulnificus Molecular Probes Antibodies, and Proteins
; FILE REFERENCE: 000432-00009
; CURRENT APPLICATION NUMBER: PCT/US98/25729
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/12467
; EARLIER FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Vibrio vulnificus
PCT-US98-25729-2

Query Match 67.3%; Score 37; DB 1; Length 410;
Best Local Similarity 75.0%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTV 9
Db 152 WWHADGTI 159

RESULT 9
PCT-US98-25729-12
; Sequence 12, Application PC/TUS9825729
; GENERAL INFORMATION:
; APPLICANT: Wright C, Anita
; APPLICANT: Powell L, Jan
; APPLICANT: Morris J, Glenn
; APPLICANT: University of Maryland Biotechnology Institute
; TITLE OF INVENTION: V.Vulnificus Molecular Probes Antibodies, and Proteins
; FILE REFERENCE: 000432-00010
; CURRENT APPLICATION NUMBER: US/09/205,283
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/12467
; EARLIER FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

Query Match 67.3%; Score 37; DB 1; Length 494;
Best Local Similarity 75.0%; Pred. No. 86;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTV 9
Db 236 WWHADGTI 243

RESULT 10
PCT-US98-25729-14
; Sequence 14, Application PC/TUS9825729
; GENERAL INFORMATION:
; APPLICANT: Wright C, Anita
; APPLICANT: Powell L, Jan
; APPLICANT: Morris J, Glenn
; APPLICANT: University of Maryland Biotechnology Institute
; TITLE OF INVENTION: V.Vulnificus Molecular Probes Antibodies, and Proteins
; FILE REFERENCE: 000432-00009
; CURRENT APPLICATION NUMBER: PCT/US98/25729
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/12467
; EARLIER FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Vibrio vulnificus
PCT-US98-25729-14

Query Match 67.3%; Score 37; DB 1; Length 410;
Best Local Similarity 75.0%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTV 9
Db 152 WWHADGTI 159

RESULT 11
US-09-205-283-2
; Sequence 2, Application US/09205283
; GENERAL INFORMATION:
; APPLICANT: Wright C, Anita
; APPLICANT: Powell L, Jan
; APPLICANT: Morris J, Glenn
; APPLICANT: University of Maryland Biotechnology Institute
; TITLE OF INVENTION: V.Vulnificus Molecular Probes Antibodies, and Proteins
; FILE REFERENCE: 000432-00010
; CURRENT APPLICATION NUMBER: US/09/205,283
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/12467
; EARLIER FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

```

; LENGTH: 410
; TYPE: PRT
; ORGANISM: Vibrio vulnificus
US-09-205-283-2

Query Match 67.3%; Score 37; DB 16; Length 410;
Best Local Similarity 75.0%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTV 9
|||
Db 152 WWHADGTT 159

RESULT 12

US-09-205-283-12
; Sequence 12, Application US/09205283

; GENERAL INFORMATION:

; APPLICANT: Wright C, Anita

; APPLICANT: Powell L, Jan

; APPLICANT: Morris J, Glenn

; APPLICANT: University of Maryland Biotechnology Institute

; TITLE OF INVENTION: V.Vulnificus Molecular Probes Antibodies, and Proteins

; FILE REFERENCE: 000432-00010

; CURRENT APPLICATION NUMBER: US/09/205,283

; CURRENT FILING DATE: 1998-12-04

; EARLIER APPLICATION NUMBER: PCT/US98/12467

; EARLIER FILING DATE: 1998-06-19

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12

; LENGTH: 494

; TYPE: PRT

; ORGANISM: Vibrio vulnificus

US-09-205-283-12

Query Match

Best Local Similarity 67.3%; Score 37; DB 16; Length 494;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTV 9
|||
Db 236 WWHADGTT 243

RESULT 13

US-09-205-283-14
; Sequence 14, Application US/09205283

; GENERAL INFORMATION:

; APPLICANT: Wright C, Anita

; APPLICANT: Powell L, Jan

; APPLICANT: Morris J, Glenn

; APPLICANT: University of Maryland Biotechnology Institute

; TITLE OF INVENTION: V.Vulnificus Molecular Probes Antibodies, and Proteins

; FILE REFERENCE: 000432-00010

; CURRENT APPLICATION NUMBER: US/09/205,283

; CURRENT FILING DATE: 1998-12-04

; EARLIER APPLICATION NUMBER: PCT/US98/12467

; EARLIER FILING DATE: 1998-06-19

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 14

; LENGTH: 410

; TYPE: PRT

; ORGANISM: Vibrio vulnificus

US-09-205-283-14

Query Match

Best Local Similarity 67.3%; Score 37; DB 16; Length 410;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTV 9
|||
Db 152 WWHADGTT 159

RESULT 14

US-09-328-352-6648

; Sequence 6648, Application US/09328352

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 6648

; LENGTH: 846

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-6648

Query Match

Best Local Similarity 67.3%; Score 37; DB 17; Length 846;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWVTADGTVN 10
|||
Db 142 AWVQADPSVN 151

RESULT 15

US-60-128-476-4008

; Sequence 4008, Application US/60128476

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA

; FILE REFERENCE: 107196.139

; CURRENT APPLICATION NUMBER: US/60/128,476

; CURRENT FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 5002

; SEQ ID NO 4008

; LENGTH: 227

; TYPE: PRT

; ORGANISM: Moraxella catarrhalis

US-60-128-476-4008

Query Match

Best Local Similarity 65.5%; Score 36; DB 19; Length 227;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 WVTADG 7
|||
Db 132 WVTADG 137

Search completed: November 13, 1999, 05:08:37
Job time: 10863 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:08:00 ; Search time 75.45 Seconds
(without alignments)
5.310 Million cell updates/sec

Title: US-08-913-430-8
Perfect score: 55
Sequence: 1 AWVTADGTVN 10

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR-60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	70.9	230	1 B8AG6	virB8 protein - Ag
2	39	70.9	492	2 E64970	hypothetical prote
3	37	67.3	140	2 E70832	hypothetical prote
4	36	65.5	724	1 RRWTD	RNA-directed RNA p
5	36	65.5	241	2 S43199	cell division cont
6	36	65.5	734	2 I38080	cell division cont
7	36	65.5	403	2 H69362	succinyl-diaminop
8	36	65.5	229	2 S30949	gene 1 protein - M
9	36	65.5	378	2 S56146	probable polysacch
10	36	65.5	4199	2 S76412	hypothetical prote
11	36	65.5	544	2 I55454	neuroglycan C prec
12	36	65.5	456	2 G71379	probable chromate
13	35	63.6	2339	2 S41121	acetyl-CoA carboxy
14	35	63.6	3124	2 A40020	collagen alpha 1(X
15	35	63.6	108	2 S37313	hlyU protein - Vib
16	35	63.6	574	2 I37426	glutamate transpor
17	35	63.6	565	2 I38399	glutamate/aspartin
18	35	63.6	572	2 A55676	excitatory amino a
19	35	63.6	574	2 JC4267	glutamate transpor
20	35	63.6	574	2 I38432	excitatory amino a
21	35	63.6	573	2 S28901	glutamate transpor
22	35	63.6	503	2 JC5078	glutamate transpor
23	35	63.6	430	2 JC5580	glutamate transpor
24	35	63.6	492	2 J01050	deoxyfructosyl-glu
25	35	63.6	417	2 A64458	branched-chain ami
26	35	63.6	101	2 E71438	probable transcrip
27	35	63.6	918	2 A41943	vacuolar membrane
28	34	61.8	257	1 B8AG55	virB8 protein - Ag
29	34	61.8	237	1 B8AG58	virB8 protein - Ag
30	34	61.8	272	2 C44816	N-acetylmuramoyl-L
31	34	61.8	317	2 B69734	CIP synthase - Met
32	34	61.8	533	2 E69154	acetyl-CoA carboxy
33	34	61.8	2233	2 S63347	acetyl-CoA carboxy
34	34	61.8	2089	2 A48757	mannopine biosynth
35	34	61.8	430	2 I39727	polysaccharide exp
36	34	61.8	377	2 S61892	probable polysacch
37	34	61.8	379	2 E64839	probable transloca
38	34	61.8	808	2 E70720	hypothetical prote
39	34	61.8	825	2 T00818	hypothetical prote

SSD1 protein - yea
interleukin-6 sign
E2 protein - human
E2 protein - human
nucleocapsid prote
acetyl-CoA carboxy

ALIGNMENTS

RESULT 1
B8AG6
virB8 protein - Agrobacterium tumefaciens plasmid pTiA6
C:Species: Agrobacterium tumefaciens
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 03-Feb-1994
C:Accession: I28621; I27127
R:Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.
J. Biol. Chem. 263, 5804-5814, 1988
A:Title: Characterization of the virB operon from an Agrobacterium tumefaciens Ti plasmid
A:Reference number: A28621; MUID:88186901
A:Accession: I28621
A:Molecule type: DNA
A:Residues: 1-230 <WAP>
A:Cross-references: GB:J03216
A:Note: This sequence was designated ORF 9 in this reference
R:Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.
J. Biol. Chem. 265, 4768, 1990
A:Reference number: A35737; MUID:90170994
A:Contents: annotation; erratum
C:Genetics:
A:Genome: plasmid
C:Superfamily: tumor-inducing plasmid pTiC58 virB8 protein

Query Match 70.9%; Score 39; DB 1; Length 230;
Best Local Similarity 66.7%; Pred. No. 3.7;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
DB 66 WVTADGTVD 74

RESULT 2
E64970
hypothetical protein b2046 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 13-Mar-1998
C:Accession: E64970
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Riley, M.;
A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E64970
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-492 <BLAT>
A:Cross-references: GB:AE000295; GB:U00096; NID:g1788354; PID:g1788359; UWGP:b2046
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: hypothetical protein b2046

Query Match 70.9%; Score 39; DB 2; Length 492;
Best Local Similarity 60.0%; Pred. No. 8.1;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWVTADGTVN 10
DB 214 AWVTADSIIN 223

RESULT 3

E70632
 hypothetical protein Rv0390 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
 C:Accession: E70632
 R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: E70632
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-140 <COL>
 A:Cross-references: GB:284725; GB:AL123456; NID:93261703; PID:e300409; PID:g1817705
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv0390

Query Match 67.3%; Score 37; DB 2; Length 140;
 Best Local Similarity 66.7%; Pred. No. 5.2;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
 | | | | |
 Db 53 WATSDGTHN 61

RESULT 4

RWQVD
 RNA-directed RNA polymerase (EC 2.7.7.48) - tobacco necrosis virus (strain D)
 N:Alternate names: RNA nucleotidyltransferase (RNA-directed); RNA replicase
 N:Contains: 22K protein
 C:Species: tobacco necrosis virus, TNV
 C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 25-Oct-1996
 C:Accession: J00368
 R:Countts, R.H.A.; Rigden, J.E.; Slabas, A.R.; Lomonosoff, G.P.; Wise, P.J.
 J. Gen. Virol. 72, 1521-1529, 1991
 A:Title: The complete sequence of tobacco necrosis virus strain D.
 A:Reference number: J00368; MUID:91311404
 A:Accession: J00368
 A:Molecule type: genomic RNA
 A:Residues: 1-724 <COL>
 A:Cross-references: GB:D00942
 A:Note: readthrough of the terminator TAG occurs between codons AAA for 202-Lys and GGA
 C:Superfamily: carnation mottle virus RNA-directed RNA polymerase; barley yellow dwarf
 C:Keywords: nucleotidyltransferase; RNA biosynthesis; RNA replication
 F:1-202/Product: 22K protein #status predicted <KPT>
 F:427-597/Domain: barley yellow dwarf virus RNA-directed RNA polymerase homology <BYD>

Query Match 65.5%; Score 36; DB 1; Length 724;
 Best Local Similarity 87.5%; Pred. No. 44;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AWVTADGT 8
 | | | | |
 Db 612 AWVTAVGT 619

RESULT 5

S43199
 cell division control protein CDC46 - human (fragment)
 N:Alternate names: S. cerevisiae minichromosome maintenance deficient 5 homolog
 C:Species: Homo sapiens (man)
 C>Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 04-Sep-1998
 C:Accession: S43199; S41623
 R:Hu, B.
 submitted to the EMBL Data Library, August 1993

A:Reference number: S43198

A:Accession: S43199
 A:Molecule type: mRNA
 A:Residues: 1-241 <HUB1>
 A:Cross-references: EMBL:X74795
 R:Hu, B.; Burkhardt, R.; Schulte, D.; Musahl, C.; Knippers, R.
 Nucleic Acids Res. 21, 5289-5293, 1993
 A:Title: The P1 family: a new class of nuclear mammalian proteins related to the yeast
 A:Reference number: S41622; MUID:94089373
 A:Accession: S41623
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-158 <HUB2>
 A:Cross-references: EMBL:X74795
 C:Genetics:
 A:Gene: GDB:MCM5; CDC46
 A:Cross-references: GDB:433799
 A:Map position: 22q13.1-22q13.2
 C:Superfamily: cell division control protein CDC47
 C:Keywords: cell division control; DNA binding; transcription regulation
 F:1-191/Domain: MCM2 core domain similarity (fragment) <MCM>

Query Match 65.5%; Score 36; DB 2; Length 241;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWVTADGTV 9
 | | | | |
 Db 84 AWVLADGGV 92

RESULT 6

I38080
 cell division control protein CDC46 - human
 N:Alternate names: S. cerevisiae minichromosome maintenance deficient 5 homolog
 C:Species: Homo sapiens (man)
 C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 04-Sep-1998
 C:Accession: I38080
 R:Hu, B.; Burkhardt, R.; Schulte, D.; Musahl, C.; Knippers, R.
 Nucleic Acids Res. 21, 5289-5293, 1993
 A:Title: The P1 family: a new class of nuclear mammalian proteins related to the yeast
 A:Reference number: I38080; MUID:94089373
 A:Accession: I38080
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-734 <RES>
 A:Cross-references: EMBL:X74795; NID:g895842; PID:g895843
 C:Genetics:
 A:Gene: GDB:MCM5; CDC46
 A:Cross-references: GDB:433799
 A:Map position: 22q13.1-22q13.2
 C:Superfamily: cell division control protein CDC47
 F:328-540/Domain: MCM2 core domain similarity <MCM>

Query Match 65.5%; Score 36; DB 2; Length 734;
 Best Local Similarity 77.8%; Pred. No. 45;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWVTADGTV 9
 | | | | |
 Db 433 AWVLADGGV 441

RESULT 7

H69362
 succinyl-diaminopimelate desuccinylase (dapE-2) homolog - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 18-Sep-1998
 C:Accession: H69362
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiaach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
A:Reference number: A69250; MUID:98049343
A:Accession: H69362
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-403 <KLE>
A:Cross-references: GB:AE001041; GB:AE000782; NID:g2689364; PID:g2649695; TIGR:AF0904
C:Superfamily: succinyl-diaminopimelate desuccinylase

Query Match 65.5%; Score 36; DB 2; Length 403;
Best Local Similarity 60.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AWTADGTVN 10
|||
Db 367 AWTADGVAH 376

RESULT 8
S30949
gene 1 protein - Mycobacterium phage L5
C:Species: Mycobacterium phage L5
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Sep-1997
C:Accession: S30949
R:Donnelly-Wu, M.K.; Jacobs Jr., W.R.; Hatfull, G.F.
Mol. Microbiol. 7, 407-417, 1993
A:Title: Superinfection immunity of mycobacteriophage L5: applications for genetic trans
A:Reference number: S30949
A:Accession: S30949
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-229 <DON>
A:Cross-references: EMBL:218946; NID:g15859; PID:e59622; PID:g579116
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992
C:Genetics:
A:Gene: 1
A:Start codon: GTG

Query Match 65.5%; Score 36; DB 2; Length 229;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 WVTADG 7
|||||
Db 220 WVTADG 225

RESULT 9
B56146
probable polysaccharide export protein precursor - Klebsiella pneumoniae
C:Species: Klebsiella pneumoniae
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 24-Oct-1998
C:Accession: B56146
R:Arakawa, Y.; Wacharotayankun, R.; Nagatsuka, T.; Ito, H.; Kato, N.; Ohta, M.
J. Bacteriol. 177, 1788-1796, 1995
A:Title: Genomic organization of the Klebsiella pneumoniae cps region responsible for se
A:Reference number: A56146; MUID:95204345
A:Accession: B56146
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <ARA>
A:Cross-references: GB:D21242; NID:g747654; PID:dl005308; PID:g747663
A:Experimental source: strain Chedid
C:Keywords: lipoprotein; polysaccharide export; transport protein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-378/Product: probable polysaccharide export protein #status predicted <MAT>

Query Match 65.5%; Score 36; DB 2; Length 378;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGT 9
|||
Db 119 WVNDSGTI 126

RESULT 10
S76412
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
C:Accession: S76412
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima,
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4199 <KAN>
A:Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PID:dl019274; PID:g165362
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 65.5%; Score 36; DB 2; Length 4199;
Best Local Similarity 55.6%; Pred. No. 2.8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
|||
Db 2251 WLAADGSLN 2259

RESULT 11
I55454
neuroglycan C precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Jul-1996
C:Accession: I55454
R:Watanabe, E.; Maeda, N.; Matsui, F.; Kushima, Y.; Noda, M.; Oohira, A.
J. Biol. Chem. 270, 26876-26882, 1995
A:Title: Neuroglycan C, a novel membrane-spanning chondroitin sulfate proteoglycan th
A:Reference number: I55454; MUID:96070781
A:Accession: I55454
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-544 <RES>
A:Cross-references: EMBL:U033553; NID:gi061328; PID:gi061329

Query Match 65.5%; Score 36; DB 2; Length 544;
Best Local Similarity 75.0%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AWTADGT 8
|||
Db 105 AWLEADGT 112

RESULT 12
G71379
probable chromate transport protein - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 28-Aug-1998 #sequence_revision 28-Aug-1998 #text_change 17-Mar-1999
C:Accession: G71379
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDevitt, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 Science 281, 375-388, 1998
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
 A:Accession number: A7150; MUID:98332770
 A:Accession: G71379
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-456 <COL>
 A:Cross-references: GB:AE001199; GB:AE000520; NID:g3322402
 A:Experimental source: strain Nichols
 A:Note: This ORF is annotated but not translated in GenBank entry AE001199, release 107
 C:Comment: This is the hypothetical translation of a sequence that was reported as a coding region of the genome of Treponema pallidum, the syphilis spirochete.
 C:Comment: This translation was produced by PIR staff from information in the GenBank accession codon. A BLAST search of the GenBank database shows that, in the majority of cases, the translation is correct.
 C:Genetics:
 A:Gene: TP0146
 C:Keywords: translational frameshift
 F:264-265/Region: plus-two translational frameshift

Query Match 65.5%; Score 36; DB 2; Length 456;
 Best Local Similarity 87.5%; Pred. No. 27;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AWWTADGT 8
 |||||
 Db 221 AWWTAMGT 228

RESULT 13
 S41121
 acetyl-CoA carboxylase (EC 6.4.1.2) - human
 C:Species: Homo sapiens (man)
 C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 05-Mar-1999
 C:Accession: S41121
 R:Ha, J.; Daniel, S.; Kong, I.-S.; Park, C.K.; Tae, H.J.; Kim, K.H.
 Eur. J. Biochem. 219, 297-306, 1994
 A:Title: Cloning of human acetyl-CoA carboxylase cDNA.
 A:Reference number: S41121; MUID:94139704
 A:Accession: S41121
 A:Molecule type: mRNA
 A:Residues: 1-2339 <HAJ>
 A:Cross-references: EMBL:X68968; NID:g452315; PID:g452316
 C:Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin-
 C:Keywords: biotin; ligase
 F:120-620/Domain: biotin carboxylase homology <BC>
 F:747-819/Domain: lipoyl/biotin-binding homology <LPB>
 F:786/Binding site: biotin (Lys) (covalent) #status predicted

Query Match 63.6%; Score 35; DB 2; Length 2339;
 Best Local Similarity 50.0%; Pred. No. 2.3e+02;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AWWTADGTVN 10
 :|||
 Db 2084 SWVVIDATIN 2093

RESULT 14
 A40020
 collagen alpha 1(XII) chain precursor - chicken
 N:Alternate names: fibrochimerin
 C:Species: Gallus gallus (chicken)
 C:Date: 28-May-1992 #sequence_revision 28-Oct-1994 #text_change 12-Feb-1999
 C:Accession: A40020; A34485; B34485; A28037; S23814; S22254; S28811
 R:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Obata, T.
 J. Cell Biol. 115, 209-221, 1991
 A:Title: The complete primary structure of type XII collagen shows a chimeric molecule with a region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.
 A:Reference number: A40020; MUID:92011862
 A:Accession: A40020

A:Molecule type: mRNA
 A:Residues: 1-3124 <YAK>
 A:Cross-references: GB:D00824; NID:g222810; PID:d1001160; PID:g222811
 A:Note: In the authors' translation residues 1216-1219 are shown after residue 1235 a
 R:Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
 J. Biol. Chem. 264, 19772-19778, 1989
 A:Title: Type XII collagen. A large multidomain molecule with partial homology to type
 A:Reference number: A34485; MUID:90062079
 A:Accession: A34485
 A:Molecule type: mRNA
 A:Residues: 2456-2978, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
 A:Cross-references: EMBL:J05137; NID:g211284; PID:g211285
 A:Accession: B34485
 A:Molecule type: protein
 A:Residues: 2772-2792; 2846-2873 <GOR2>
 R:Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
 Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
 A:Title: Type XII collagen: distinct extracellular matrix component discovered by cDN
 A:Reference number: A28037; MUID:87317590
 A:Accession: A28037
 A:Molecule type: mRNA
 A:Residues: 2960-2976, 'F', 2978-3074, 'AG', <GOR3>
 A:Cross-references: EMBL:M17375; NID:g211649; PID:g211650
 A:Note: this sequence has been revised in reference A34485
 R:Koch, M.; Bernasconi, C.; Chiquet, M.
 Eur. J. Biochem. 207, 847-856, 1992
 A:Title: A major oligomeric fibroblast proteoglycan identified as a novel large form
 A:Reference number: S23814; MUID:92362621
 A:Accession: S23814
 A:Molecule type: protein
 A:Residues: 'X', 1333, 'Q', 1335-1347; 1914-1928; 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-251
 R:Dublet, B.; van der Rest, M.
 J. Biol. Chem. 262, 17724-17727, 1987
 A:Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of peps
 A:Reference number: S22254; MUID:88087065
 A:Accession: S22254
 A:Molecule type: protein
 A:Residues: 2831-2832, 'T', 2834, 'R', 2836-2843; 3002-3014 <DUB>
 R:Trueb, J.; Trueb, B.
 Biochim. Biophys. Acta 1171, 97-98, 1992
 A:Title: The two splice variants of collagen XII share a common 5' end.
 A:Reference number: S28811; MUID:93042014
 A:Accession: S28811
 A:Molecule type: mRNA
 A:Residues: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>
 C:Genetics:
 A:Introns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
 C:Superfamily: collagen alpha 1(XII) chain; fibronectin type III repeat homology; von
 C:Keywords: alternative splicing; cell binding; coiled coil; connective tissue; disul
 F:1-23/Domain: signal sequence status predicted <SIG>
 F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
 F:24, 1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predict
 F:24-114/Domain: IIIA <IIIA>
 F:24-105/Domain: fibronectin type III repeat homology <FN3A>
 F:137-301/Domain: von Willebrand factor type A repeat homology <VWA1>
 F:332-425/Domain: IIIB <IIIB>
 F:332-414/Domain: fibronectin type III repeat homology <FN3B>
 F:437-601/Domain: von Willebrand factor type A repeat homology <VWA2>
 F:629-1178/Domain: IIIC <IIIC>
 F:630-711/Domain: fibronectin type III repeat homology <FN3C>
 F:721-802/Domain: fibronectin type III repeat homology <FN3D>
 F:812-895/Domain: fibronectin type III repeat homology <FN3E>
 F:905-986/Domain: fibronectin type III repeat homology <FN3F>
 F:995-1076/Domain: fibronectin type III repeat homology <FN3G>
 F:1086-1169/Domain: fibronectin type III repeat homology <FN3H>
 F:1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>
 F:1384-2295/Domain: IIID <IIID>
 F:1384-1465/Domain: fibronectin type III repeat homology <FN3I>
 F:1474-1557/Domain: fibronectin type III repeat homology <FN3J>
 F:1566-1647/Domain: fibronectin type III repeat homology <FN3K>
 F:1655-1738/Domain: fibronectin type III repeat homology <FN3L>

F:1756-1838/Domain: fibronectin type III repeat homology <FN3M>
F:1847-1928/Domain: fibronectin type III repeat homology <FN3N>
F:1937-2019/Domain: fibronectin type III repeat homology <FN3O>
F:2028-2110/Domain: fibronectin type III repeat homology <FN3P>
F:2119-2199/Domain: fibronectin type III repeat homology <FN3Q>
F:2207-2294/Domain: fibronectin type III repeat homology <FN3R>
F:2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>
F:2438-2440/Region: cell attachment #status predicted
F:2509-2750/Domain: IXP, homologous to NC4 domain of type IX collagen <IXP>
F:2751-2902/Domain: collagenous COL2 #status predicted <COL2>
F:2899-2901/Region: cell attachment (R-G-D) motif
F:2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>
F:2946-3048/Domain: collagenous COL1 #status predicted <COL1>
F:3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>
F:32,1006,1032,1044,1512,1767,2210,2273,2532,2683/Binding site: carbohydrate (Asn) (cova
F:2780,2789,2836,2842,2860,2866,2869,3004,3007/Modified site: hydroxyproline (Pro) #stat

Query Match 63.6%; Score 35; DB 2; Length 3124;
Best Local Similarity 60.0%; Pred. No. 3.le+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AWTATDGTVN 10
 :| | | | |
Db 925 SWTPAGIVN 934

RESULT 15
S37313
hlyU protein - Vibrio cholerae
C:Species: Vibrio cholerae
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Aug-1998
C:Accession: S37313
R:Williams, S.G.; Attridge, S.R.; Manning, P.A.
Mol. Microbiol. 9, 751-760, 1993
A:Title: The transcriptional activator HlyU of Vibrio cholerae: nucleotide sequence and
A:Reference number: S37312; MUID:94049116
A:Accession: S37313
A:Molecule type: DNA
A:Residues: 1-108 <WIL>
A:Cross-references: EMBL:X66866; NID:g403330; PID:g403332
C:Genetics:
A:Gene: hlyU
C:Superfamily: arsenical resistance operon repressor

Query Match 63.6%; Score 35; DB 2; Length 108;
Best Local Similarity 60.0%; Pred. No. 9.4;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AWTATDGTVN 10
 || | | | |
Db 66 AWRDRDGLVN 75

Search completed: November 13, 1999, 12:08:01
Job time: 2073 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:33:37 ; Search time 51.07 Seconds
(without alignments)
5.535 Million cell updates/sec

Title: US-08-913-430-8
Perfect score: 55
Sequence: 1 AWTADGTVN 10

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	70.9	230	1	VIB8_AGR6
2	39	70.9	492	1	WZXC_ECOLI
3	37	67.3	467	1	VE2_HPV24
4	36	65.5	724	1	RRPO_TNVD
5	36	65.5	229	1	VG01_BPML5
6	36	65.5	269	1	VG69_BPMD2
7	36	65.5	378	1	YC04_KLEPN
8	35	63.6	3124	1	CAIC_CHICK
9	35	63.6	2483	1	COA2_HUMAN
10	35	63.6	503	1	EAT1_CAMEL
11	35	63.6	574	1	EAT2_HUMAN
12	35	63.6	572	1	EAT2_MOUSE
13	35	63.6	573	1	EAT2_RAT
14	35	63.6	108	1	HLXU_VIBCH
15	35	63.6	430	1	MAS1_AGRRA
16	35	63.6	918	1	PEP3_YEAST
17	34	61.8	377	1	AMSH_ERWAM
18	34	61.8	2279	1	COAC_SCHPO
19	34	61.8	2233	1	COAC_YEAST
20	34	61.8	519	1	CP11_SHEEP
21	34	61.8	272	1	CWLA_BACSU
22	34	61.8	633	1	DNAB_BRAJA
23	34	61.8	918	1	IL6B_RAT
24	34	61.8	430	1	MAS1_AGRRC
25	34	61.8	808	1	SEA2_MYCTU
26	34	61.8	1038	1	SOG_DROME
27	34	61.8	370	1	SOP2_HUMAN
28	34	61.8	1250	1	SSD1_YEAST
29	34	61.8	237	1	VIB8_AGR5
30	34	61.8	230	1	VIB8_AGR9
31	34	61.8	379	1	YCC2_ECOLI
32	33	60.0	372	1	AR41_HUMAN
33	33	60.0	319	1	BGAM_LEULA
34	33	60.0	513	1	C861_ARATH
35	33	60.0	290	1	CHIT_SACER
36	33	60.0	2346	1	COAL_HUMAN
37	33	60.0	2345	1	COAC_RAT
38	33	60.0	2346	1	COAC_SHEEP
39	33	60.0	252	1	DHSB_BACSU
40	33	60.0	1012	1	DPOG_PICPA
41	33	60.0	2273	1	HFAL_YEAST
42	33	60.0	382	1	NCAP_CVCAE
43	33	60.0	382	1	NCAP_CVPPS

ALIGNMENTS

```
RESULT 1
VIB8_AGR6 382 1 NCAP_CVPPU P04134 porcine tra
45 33 60.0 514 1 VE2_HPV5B P26545 human papil

AC P09781; STANDARD; PRT; 230 AA.
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE VIRB8 PROTEIN.
GN VIRB8.
OS AGROBACTERIUM TUMEFACIENS.
OG PLASMID PTIA6.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88186901.
RA WARD J.E., AKIYOSHI D.E., REGIER D., DATTA A., GORDON M.P.,
RA NESTER E.W.;
RT "Characterization of the virB operon from an Agrobacterium
RT tumefaciens Ti plasmid."
RL J. BIOL. CHEM. 263:5804-5814(1988).
RN [2]
RP REVISIONS.
RX MEDLINE; 90170994.
RA WARD J.E., AKIYOSHI D.E., REGIER D., DATTA A., GORDON M.P.,
RA NESTER E.W.;
RL J. BIOL. CHEM. 265:4768-4768(1990).
CC -1- FUNCTION: VIRB PROTEINS ARE SUGGESTED TO ACT AT THE BACTERIAL
CC SURFACE AND THERE PLAY AN IMPORTANT ROLE IN DIRECTING T-DNA
CC TRANSFER TO PLANT CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J03216; ; NOT_ANNOTATED_CDS.
CC FIR: I28621; B8AGA6.
CC KW CROWN GALL TUMOR; PLASMID.
CC SQ SEQUENCE 230 AA; 25382 MW; 448A66C CRC32;

Query Match 70.9%; Score 39; DB 1; Length 230;
Best Local Similarity 66.7%; Pred. No. 3.6;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
Db 66 WIRADGTVD 74

RESULT 2
WZXC_ECOLI STANDARD; PRT; 492 AA.
ID WZXC_ECOLI 008002;
AC P77377; 008002; 008002;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE LIPOPOLYSACCHARIDE BIOSYNTHESIS PROTEIN WZXC.
GN WZXC OR WZX.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
```

```
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 96326333.
RA STEVENSON G., HOBBS M., ANDRIANPOULOS K., REEVES P.;
RT "Organization of the Escherichia coli K-12 gene cluster responsible
RT for production of the extracellular polysaccharide colanic acid.";
RN J. BACTERIOL. 178:4885-4893(1996).
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN SCIENCE 277:1453-1474(1997).
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 97251358.
RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T., ISONO K.,
RA KASAI H., KIMURA S., KITARAWA M., KITAGAWA K., MAKINO K., MIKI T.,
RA MIYOSHI K., MORI H., MORI T., MOTOMURA K., NAKADE S., NAKAMURA Y.,
RA NASHIMOTO H., NISHIO Y., OSHIMA T., SAITO N., SAMPEI G., SEKI Y.,
RA SIVASUNDARAM S., TAGAMI H., TAKEDA J., TAKEMOTO K., WADA C.,
RA YAMAMOTO Y., HORIUCHI T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RN DNA RES. 3:379-392(1996).
CC -1- FUNCTION: PROBABLE EXPORT PROTEIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE ANSL/EXOT/GUMJ FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U38473; G1407619; -.
DR EMBL; AE000295; G1788359; -.
DR EMBL; D90842; G1736748; -.
DR EMBL; D90843; G1736752; -.
DR ECOGENE; EGI3576; WZXC.
KW LIPOLYSACCHARIDE BIOSYNTHESIS; TRANSPORT; TRANSMEMBRANE.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 43 63 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 105 125 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 289 309 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 365 385 POTENTIAL.
FT TRANSMEM 387 407 POTENTIAL.
FT TRANSMEM 418 438 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
SQ SEQUENCE 492 AA; 53692 MW; 5ACD7EF4 CRC32;
```

```
Query Match 70.9%; Score 39; DB 1; Length 492;
Best Local Similarity 60.08; Pred. No. 7.7;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AWTADGTN 10
Db 214 AWTADSIIN 223
```

```
RESULT 3
VE2_HPV24 STANDARD; PRT; 467 AA.
AC P50770;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE REGULATORY PROTEIN E2.
GN E2.
OS HUMAN PAPILLOMAVIRUS TYPE 24.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
[1]
RN RP SEQUENCE FROM N.A.
RA DELIUS H.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U31782; G1020198; -.
DR PFAM; PF00508; E2.N; 1.
DR PFAM; PF00511; E2.C; 1.
DR HSP; P03122; 2BOP.
SQ SEQUENCE 467 AA; 52856 MW; 8E6D9A26 CRC32;

Query Match 67.3%; Score 37; DB 1; Length 467;
Best Local Similarity 75.0%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AWTADGT 8
Db 419 SWAADGT 426

RESULT 4
RPQ_TNVD STANDARD; PRT; 724 AA.
AC P27209;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) [CONTAINS: 22 KD PROTEIN].
OS TOBACCO NECROSIS VIRUS (STRAIN D) (TNV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NECROVIRUS.
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE; 91311404.
RA COUTTS R.H.A., RIGDEN J.E., SLABAS A.R., LOMONOSOFF G.P., WISE P.J.;
RT "The complete nucleotide sequence of tobacco necrosis virus strain
RT D.";
RL J. GEN. VIROL. 72:1521-1529(1991).
[2]
RN RP REVISIONS.
RA COUTTS R.H.A.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- READTHROUGH OF A TERMINATOR CODON OCCURS BETWEEN CODONS FOR
CC LYS-202 AND GLY-203.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D00942; G1638814; -.
```

```

DR EMBL; D00942; G1638815; ALT_TERM.
DR PIR; J00368; RWOQD.
KW TRANSFERASE; RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 202 22 KD PROTEIN.
SQ SEQUENCE 724 AA; 82245 MW; 3349FA32 CRC32;

Query Match 65.5%; Score 36; DB 1; Length 724;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AWTADGT 8
   |||||
Db 612 AWTAVGT 619

RESULT 5
VG01_BPML5 STANDARD; PRT; 229 AA.
AC Q05218;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE GENE 1 PROTEIN (GPI).
GN 1.
OS MYCOBACTERIOPHAGE L5.
OC VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93211282.
RA HATFULL G.F., SARKIS G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
RT a phage system for mycobacterial genetics.";
RL MOL. MICROBIOL. 7:395-405(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z18946; E1192380;
DR PIR; S30949; S30949.
SQ SEQUENCE 229 AA; 26475 MW; 2464C60F CRC32;

Query Match 65.5%; Score 36; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 WVTADG 7
   |||||
Db 220 WVTADG 225

RESULT 6
VG69_BPMD2 STANDARD; PRT; 269 AA.
AC O84262;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENE 69 PROTEIN (GP69).
GN 69.
OS MYCOBACTERIOPHAGE D29.
OC VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98300335.
RA FORD M.E., SARKIS G.J., BELANGER A.E., HENDRIX R.W., HATFULL G.F.;
RT "Genome structure of mycobacteriophage D29: implications for phage
evolution.";
J. MOL. BIOL. 279:143-164(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF022214; G3172318;
DR PIR; S30949; S30949.
SQ SEQUENCE 269 AA; 30781 MW; 695A4101 CRC32;

Query Match 65.5%; Score 36; DB 1; Length 269;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGT 8
   |||||
Db 131 WVTADGT 137

RESULT 7
YC04_KLEPN STANDARD; PRT; 378 AA.
AC Q48450;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PUTATIVE CAPSULE POLYSACCHARIDE EXPORT PROTEIN PRECURSOR (ORE4).
OS KLEBSIELLA PNEUMONIAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC KLEBSIELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-CHIEDID;
RX MEDLINE; 95204345.
RA ARAKAWA Y., WACHAROTAYANKUN R., NAGATSUKA T., ITO H., KATO N.,
RA OHTA M.;
RT "Genomic organization of the Klebsiella pneumoniae cps region
RT responsible for serotype K2 capsular polysaccharide synthesis in the
RT virulent strain Chediak.";
RL J. BACTERIOL. 177:1788-1796(1995).
CC 1- FUNCTION: MAY BE INVOLVED IN POLYSACCHARIDE TRANSPORT.
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY).
CC 1- SIMILARITY: BELONGS TO THE BEXD/CTRA/VEXA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D21242; G747663;
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW HYPOTHETICAL PROTEIN; POLYSACCHARIDE TRANSPORT; TRANSPORT;
KW OUTER MEMBRANE; TRANSMEMBRANE; LIPOPROTEIN; PORIN; SIGNAL.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 378 PUTATIVE CAPSULE POLYSACCHARIDE EXPORT
   PROTEIN.
FT LIPID 21 21 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 378 AA; 41497 MW; 22FD42BD CRC32;

Query Match 65.5%; Score 36; DB 1; Length 378;
Best Local Similarity 62.5%; Pred. No. 21;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```


FT CARBOHYD 1032 1032 POTENTIAL.
FT CARBOHYD 1044 1044 POTENTIAL.
FT CARBOHYD 1512 1512 POTENTIAL.
FT CARBOHYD 1767 1767 POTENTIAL.
FT CARBOHYD 2210 2210 POTENTIAL.
FT CARBOHYD 2273 2273 POTENTIAL.
FT CARBOHYD 2532 2532 POTENTIAL.
FT CARBOHYD 2683 2683 POTENTIAL.
FT CARBOHYD 2683 2683 MISSING (IN SHORT FORM).
FT VARSPLIC 25 1188 T -> S (IN REF. 4).
FT CONFLICT 1258 1258 D -> E (IN REF. 4).
FT CONFLICT 1264 1264 P -> A (IN REF. 2).
FT CONFLICT 2759 2759 L -> F (IN REF. 2).
FT CONFLICT 2803 2803 V -> F (IN REF. 2).
FT CONFLICT 2977 2977 QP -> AG (IN REF. 3).
FT CONFLICT 3075 3076
SQ SEQUENCE 3124 AA; 340578 MW; 59E3DB42 CRC32;

Query Match 63.6%; Score 35; DB 1; Length 3124;
Best Local Similarity 60.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 AWTADGTGVN 10
Db 925 SWTPAGTVN 934

RESULT 9
COA2_HUMAN
ID COA2_HUMAN STANDARD; PRT; 2483 AA.
AC 000763; Q16852;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ACETYL-COA CARBOXYLASE 2 (EC 6.4.1.2) (ACC-BETA) [CONTAINS: BIOTIN
CARBOXYLASE (EC 6.3.4.14)].
GN ACACB OR ACC2 OR ACCB.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER.
RX MEDLINE; 97256787.
RA ABU-ELHEIGA L., ALMARZA-ORTEGA D.B., BALDINI A., WAKIL S.J.;
RT "Human acetyl-CoA carboxylase 2. Molecular cloning, characterization,
J. BIOL. CHEM. 272:10669-10677(1997).
[2]
RP SEQUENCE OF 1349-2134 FROM N.A.
RC TISSUE=ADIPOSE TISSUE;
RX MEDLINE; 96265061.
RA WIDMER J., FASSIHI K.S., SCHLICHTER S.C., WHEELER K.S., CRUTE B.E.,
RA KING N., NUTILE-MCENEMY N., NOLL W.W., DANIEL S., HA J., KIM K.-H.,
RA WITTEBS L.A.;
RT "Identification of a second human acetyl-CoA carboxylase gene.";
RL BIOCHEM. J. 316:915-922(1996).
CC -1- FUNCTION: ACC-BETA MAY BE INVOLVED IN THE PROVISION OF MALONYL-COA
OR IN THE REGULATION OF FATTY ACID OXIDATION, RATHER THAN FATTY
ACID BIOSYNTHESIS. THIS PROTEIN CARRIES THREE CATALYTIC FUNCTIONS:
BIOTIN CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE, AND
CARBOXYLTRANSFERASE.
CC -1- CATALYTIC ACTIVITY: ATP + ACETYL-COA + HCO(3)(-) - ADP + PHOSPHATE
+ MALONYL-COA.
CC -1- COFACTOR: BIOTIN.
CC -1- PATHWAY: FIRST STEP (RATE LIMITING) IN LONG-CHAIN FATTY ACID
SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; MAY ASSOCIATE WITH MEMBRANES.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE HEART, SKELETAL
MUSCLES AND LIVER.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS ARE PRODUCED BY ALTERNATIVE
SPLICING OF THE SAME GENE.
CC -1- SIMILARITY: PARTIAL TO CARBAMOYL PHOSPHATE SYNTHETASES.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U89344; G2138330; -;
DR EMBL; U34591; G1399290; -;
DR MIM; 601557; -;
DR PROSITE; PS00188; BIOTIN; 1.
DR PROSITE; PS00866; CPSASE_1; 1.
DR PROSITE; PS00867; CPSASE_2; 1.
DR PFAM; PF00289; CPSase_L_chain; 1.
DR PFAM; PF00364; biotin_req_enzy; 1.
DR PFAM; PF01039; Carboxyl_crans; 1.
DR HSSP; P24182; IBNC.
KW FATTY ACID BIOSYNTHESIS; BIOTIN; LIGASE; MULTIFUNCTIONAL ENZYME;
KW PHOSPHORYLATION; MULTIGENE FAMILY; ALTERNATIVE SPLICING.
FT NP_BIND 458 463 ATP (POTENTIAL).
FT ACT_SITE 584 584 BY SIMILARITY.
FT BINDING 927 927 BIOTIN.
FT DOMAIN 2095 2124 COENZYME A-BINDING (POTENTIAL).
FT MOD_RES 219 219 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 221 221 PHOSPHORYLATION (BY SIMILARITY).
FT VARSPLIC 1115 1215 MISSING (IN SHORT FORM).
FT CONFLICT 1365 1365 S -> N (IN REF. 2).
FT CONFLICT 1375 1375 T -> S (IN REF. 2).
FT CONFLICT 1408 1408 D -> G (IN REF. 2).
FT CONFLICT 1421 1421 P -> T (IN REF. 2).
FT CONFLICT 1450 1450 V -> M (IN REF. 2).
FT CONFLICT 1487 1487 P -> T (IN REF. 2).
FT CONFLICT 1555 1558 EGRY -> KEGV (IN REF. 2).
FT CONFLICT 1609 1609 N -> S (IN REF. 2).
FT CONFLICT 1630 1630 N -> F (IN REF. 2).
FT CONFLICT 1844 1844 A -> P (IN REF. 2).
FT CONFLICT 1917 1918 MI -> IM (IN REF. 2).
SQ SEQUENCE 2483 AA; 279690 MW; 87254BA9 CRC32;

Query Match 63.6%; Score 35; DB 1; Length 2483;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 AWTADGTGVN 10
Db 2228 SWVVIDATIN 2237

RESULT 10
EATL_CAEEL
ID EATL_CAEEL STANDARD; PRT; 503 AA.
AC Q10901; P90798; Q17920;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE EXCITATORY AMINO ACID TRANSPORTER (SODIUM-DEPENDENT GLUTAMATE/
ASPARTATE TRANSPORTER).
GN GLT-1 OR C12D12.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 97039669.
RA RADICE A.D., LUSTIGMAN S.;
RT "Cloning and characterization of cDNAs encoding putative glutamate
RT transporters from Caenorhabditis elegans and Onchocerca volvulus.";
RL MOL. BIOCHEM. PARASITOL. 80:41-53(1996).
RN [2]

FT TRANSMEM 121 POTENTIAL.
FT DOMAIN 143 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 240 POTENTIAL.
FT TRANSMEM 280 POTENTIAL.
FT TRANSMEM 317 POTENTIAL.
FT TRANSMEM 406 POTENTIAL.
FT TRANSMEM 437 POTENTIAL.
FT CARBOHYD 206 POTENTIAL.
FT CARBOHYD 216 POTENTIAL.
FT CARBOHYD 19 H -> P (IN REF. 1).
FT CONFLICT 19 E -> G (IN REF. 3).
FT CONFLICT 27 T -> Q (IN REF. 2).
FT CONFLICT 50 A -> S (IN REF. 2).
FT CONFLICT 58 A -> G (IN REF. 3).
FT CONFLICT 141 P -> A (IN REF. 3).
FT CONFLICT 155 V -> E (IN REF. 1).
FT CONFLICT 211 GIA -> AIP (IN REF. 3).
FT CONFLICT 253 AKLWDFENLNEIVKLVIMWYSP -> GQADGGFLQH
FT CONFLICT 263 FEROCNEVSDHSHVLS (IN REF. 3).
FT CONFLICT 347 F -> L (IN REF. 1).
FT CONFLICT 539 Y -> F (IN REF. 3).
FT CONFLICT 556 A -> G (IN REF. 3).
FT CONFLICT 563 CSV -> RVL (IN REF. 3).
FT CONFLICT 570 W -> G (IN REF. 3).
SQ SEQUENCE 574 AA; 62104 MW; D430FF87 CRC32;

Query Match 63.6%; Score 35; DB 1; Length 574;
Best Local Similarity 60.0%; Pred. No. 48;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AWVTADGTGN 10
Db 354 AWITALGTAS 363

RESULT 12

ID EAT2_MOUSE STANDARD; PRT; 572 AA.
AC P43006; Q35877;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE EXCITATORY AMINO ACID TRANSPORTER 2 (SODIUM-DEPENDENT
DE GLUTAMATE/ASPARTATE TRANSPORTER 2).
GN SLC1A2 OR EAAT2 OR GLT1.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=C57BL; TISSUE=BRAIN;
RX MEDLINE; 95213010.
RA KIRSCHNER M.A., COPLAND N.G., GILBERT D.J., JENKINS N.A.,
RA AMARA S.G.;
RT "Mouse excitatory amino acid transporter EAAT2: Isolation,
RT characterization, and proximity to neuroexcitability loci on mouse
RT chromosome 2.";
RL GENOMICS 24:218-224(1994).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=JCL;ICR; TISSUE=CEREBELLUM;
RX MEDLINE; 95284091.
RA MUKAINAKA Y., TANAKA K., HAGIWARA T., WADA K.;
RT "Molecular cloning of two glutamate transporter subtypes from mouse
RT brain.";
RL BIOCHIM. BIOPHYS. ACTA 1244:233-237(1995).
RN [3]
RP SEQUENCE FROM N.A.
RP STRAIN=C57BL/6; TISSUE=BRAIN;
RX MEDLINE; 96032356.
RA SUTHERLAND M.L., DELANEY T.A., NOBELS J.L.;
RT "Molecular characterization of a high-affinity mouse glutamate

transporter.";
RL GENE 162:271-274(1995).
RN [4]
RP SEQUENCE FROM N.A.
RP TISSUE=BRAIN;
RA PENG J.-B., GUO L.-H.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP SEQUENCE FROM N.A.
RP STRAIN=JCL;ICR; TISSUE=BRAIN;
RX MEDLINE; 98039013.
RA UTSUNOMIYA-TATE N., ENDOU H., KANAI Y.;
RT "Tissue specific variants of glutamate transporter GLT-1.";
RL FEBS LETT. 416:312-316(1997).
CC -1- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
CC ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
CC RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
CC AS A SYMPORT BY CO-TRANSPORTING SODIUM.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- PTM: GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

EMBL; U11763; G607866; -
EMBL; D43796; G984638; -
EMBL; U24699; G882343; -
EMBL; U75372; G2459554; -
EMBL; AB007810; D1024675; -
DR MGD; MGI:101931; SLC1A2.
DR PROSITE; PS00713; NA:DICARBOXYL_SYM_1; 1.
DR PROSITE; PS00714; NA:DICARBOXYL_SYM_2; 1.
DR PFAM; PF00375; SDF; 1.
KW TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN; SYMPORT; MULTIGENE FAMILY.
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 64 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 121 142 POTENTIAL.
FT DOMAIN 143 238. EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 239 300 POTENTIAL.
FT TRANSMEM 279 300 POTENTIAL.
FT TRANSMEM 316 338 POTENTIAL.
FT TRANSMEM 405 429 POTENTIAL.
FT TRANSMEM 436 458 POTENTIAL.
FT CARBOHYD 205 205 POTENTIAL.
FT CARBOHYD 215 215 POTENTIAL.
FT CONFLICT 26 26 D -> E (IN REF. 3).
FT CONFLICT 62 62 G -> R (IN REF. 3).
FT CONFLICT 112 112 A -> V (IN REF. 3).
FT CONFLICT 454 454 T -> I (IN REF. 4).
FT CONFLICT 525 525 K -> L (IN REF. 4).
FT CONFLICT 572 572 K -> EFD (IN REF. 3).
SQ SEQUENCE 572 AA; 62030 MW; D0F3C6C4 CRC32;

Query Match 63.6%; Score 35; DB 1; Length 572;
Best Local Similarity 60.0%; Pred. No. 47;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AWVTADGTGN 10
Db 353 AWITALGTAS 362

RESULT 13

EAT2_RAT
ID EAT2_RAT STANDARD; PRT; 573 AA.
AC P31596;
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE EXCITATORY AMINO ACID TRANSPORTER 2 (SODIUM-DEPENDENT
DE GLUTAMATE/ASPARTATE TRANSPORTER 2) (GLUT-R) (GLT-1).
GN SLC1A2 OR EAAT2 OR GLT1.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC KODONTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE: 93078876.
RA PINES G., DANBOLT N.C., BJORAS M., ZHANG Y., BENDAHAN A., EIDE L.,
RA KOESELL H., STORM-WATHISEN J., SEESEB E., KANNER B.I.;
RT "Cloning and expression of a rat brain L-glutamate transporter.";
RL NATURE 360:464-467(1992).
RN [2]
RP ERRATUM.
RA PINES G., DANBOLT N.C., BJORAS M., ZHANG Y., BENDAHAN A., EIDE L.,
RA KOESELL H., STORM-WATHISEN J., SEESEB E., KANNER B.I.;
RL NATURE 360:768-768(1992).
RN [3]
RP REVISIONS TO 260-289.
RX MEDLINE: 93292659.
RA KANNER B.I.;
RT "Glutamate transporters from brain. A novel neurotransmitter
RT transporter family.";
RL FEBS LETT. 325:95-99(1993).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=FOREBRAIN;
RA ROGINSKI R.S., CHOUDHURY K., MEINERS S., MARONE M., BASMA A.N.,
RA GELLER H.M.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP MUTAGENESIS OF LYS-298 AND HIS-326.
RX MEDLINE: 94308247.
RA ZHANG Y., PINES G., KANNER B.I.;
RT "Histidine 326 is critical for the function of GLT-1, a (Na⁺ + K⁺)-
RT coupled glutamate transporter from rat brain.";
RL J. BIOL. CHEM. 269:19573-19577(1994).
CC -!- FUNCTION: TRANSPORTS L-GLUTAMATE AND ALSO L- AND D-ASPARTATE.
CC ESSENTIAL FOR TERMINATING THE POSTSYNAPTIC ACTION OF GLUTAMATE BY
CC RAPIDLY REMOVING RELEASED GLUTAMATE FROM THE SYNAPTIC CLEFT. ACTS
CC AS A SYMPORT BY CO-TRANSPORTING SODIUM.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: LOCALIZED IN BRAIN AND IS HIGHLY ENRICHED IN
CC THE PURKINJE CELL LAYER IN CEREBELLUM.
CC -!- PTM: GLYCOSYLATED.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X67857; G56263; ALT_SEQ.
CC EMBL: U15098; G705398;
CC EMBL: U15098; G705397; ALT_INIT.
CC PROSITE: PS00713; NA_DICARBOXYL_SYMPT_1; 1.
CC PROSITE: PS00714; NA_DICARBOXYL_SYMPT_2; 1.
CC PFAM: PF00375; SDF; 1.
CC TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN; SYMPORT; MULTIGENE FAMILY.
FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 64
POTENTIAL.

Query Match 63.6%; Score 35; DB 1; Length 573;
Best Local Similarity 60.0%; Pred. No. 47;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 AWTADGTVN 10
DQ 353 AWITAGTAS 362
||:|||||:
ID HLYU_VIBCH STANDARD; PRT; 108 AA.
AC P32695;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TRANSCRIPTIONAL ACTIVATOR HLYU.
GN HLYU.
OS VIBRIO CHOLERAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIONACEAE; VIBRIO.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLASSICAL 01 / Z17561;
RX MEDLINE: 94049116.
RA WILLIAMS S.G., ATTRIDGE S.R., MANNING P.A.;
RT "The transcriptional activator HlyU of *Vibrio cholerae*: nucleotide
RT sequence and role in virulence gene expression.";
RL MOL. MICROBIOL. 9:751-760(1993).
CC -!- FUNCTION: UPREGULATES THE EXPRESSION OF THE HEMOLYSIN GENE, HLYA,
CC AND MAY PROMOTE EXPRESSION OF OTHER VIRULENCE DETERMINANTS IN
CC VIVO. IT MAY HAVE BOTH POSITIVE AND NEGATIVE REGULATOR ACTIVITIES.
CC -!- SIMILARITY: BELONGS TO THE ARSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X66866; G403332;
CC PFAM: PF01022; HTH_ARSR_family; 1.
CC HSP; P30340; 1SMT.
CC TRANSSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING.
FT DNA_BIND 47 66 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 108 AA; 12293 MW; 624FDC6D CRC32;

Query Match 63.6%; Score 35; DB 1; Length 108;
Best Local Similarity 60.0%; Pred. No. 9;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 AWTADGTVN 10

```
Db 66 AWRDRGLVN 75

RESULT 15
MASI_AGRRA STANDARD; PRT; 430 AA.
AC P27874;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE AGROPINE SYNTHESIS REDUCTASE (EC 1.-.-.-).
GN MASI.
OS AGROBACTERIUM RHIZOGENES.
OG PLASMID PRIA4B.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A4;
RX MEDLINE; 91239681.
RA BOUCHEZ D., TOURNEUR J.;
RT "Organization of the agropine synthesis region of the T-DNA of the R1
RL plasmid from Agrobacterium rhizogenes.";
PLASMID 25:27-39(1991).
CC -!- FUNCTION: REDUCES DEOXY-FRUCTOSYL-GLUTAMINE TO MANNOPINE.
CC -!- PATHWAY: AGROPINE / MANNOPINE SYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X51338; G407084; -
CC PIR; JQ1050; JQ1050.
CC PROSITE; PS00061; ADH_SHORT; 1.
CC PFAM; PF00106; adh_short; 1.
CC PLASMID; OXIDOREDUCTASE.
KW NP_BIND 203 227 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 346 346 BY SIMILARITY.
SQ SEQUENCE 430 AA: 47721 MW; A8875320 CRC32;

Query Match 63.6%; Score 35; DB 1; Length 430;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 9; Conservative 0; Mismatches 1; Indels 8; Caps 1;

QY 1 AWTFA-----DGTVN 10
Db 262 AWTAAVEKFGRIDGLVN 279

Search completed: November 13, 1999, 10:33:38
Job time: 5188 sec
```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:29 ; Search time 139.86 Seconds
(without alignments)
4.400 Million cell updates/sec

Title: US-08-913-430-8
Perfect score: 55
Sequence: 1 AWTADGTVN 10

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database : SPTREMBL_10:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	72.7	357	5 Q17682	Q17682 caenorhabdi
2	40	72.7	2761	5 Q19522	Q19522 caenorhabdi
3	39	70.9	230	2 Q52608	Q52608 agrobacteri
4	38.5	70.0	427	10 P93185	P93185 hordeum vul
5	38.5	70.0	377	10 P93186	P93186 hordeum vul
6	38.5	70.0	427	10 P93187	P93187 hordeum vul
7	38	69.1	438	2 Q54823	Q54823 streptomyce
8	37	67.3	140	2 P95198	P95198 mycobacteri
9	36	65.5	403	1 Q29358	Q29358 archaeoglob
10	36	65.5	99	2 Q55218	Q55218 streptomyce
11	36	65.5	333	2 Q33560	Q33560 rhodobacter
12	36	65.5	2314	2 Q69822	Q69822 streptomyce
13	36	65.5	4199	2 P74440	P74440 synecocyst
14	36	65.5	544	11 Q62831	Q62831 rattus norv
15	36	65.5	725	12 Q41347	Q41347 tobacco nec
16	35	63.6	417	1 Q58662	Q58662 methanococc
17	35	63.6	529	2 Q86946	Q86946 thermotoga
18	35	63.6	237	2 Q66288	Q66288 agrobacteri
19	35	63.6	341	2 Q921V2	Q921V2 thermotoga
20	35	63.6	592	2 Q92PD2	Q92PD2 thermotoga
21	35	63.6	815	3 Q94598	Q94598 schizosacch
22	35	63.6	101	10 Q23546	Q23546 arabidopsis
23	35	63.6	569	11 Q54686	Q54686 mus musculu
24	35	63.6	579	13 Q57322	Q57322 ambystoma t
25	35	63.6	581	13 Q57323	Q57323 ambystoma t
26	34	61.8	533	1 Q26519	Q26519 methanobact
27	34	61.8	317	2 Q34391	Q34391 bacillus su
28	34	61.8	631	2 Q05700	Q05700 rhodopsendo
29	34	61.8	286	2 Q87816	Q87816 alcaligenes

ALIGNMENTS

RESULT 1

Q17682 PRELIMINARY; PRT; 357 AA.
 ID Q17682
 AC Q17682;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE SIMILAR TO PROTEIN KINASES.
 GN C05H8.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SNALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 Nature 0:0-0(0).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA DING H.;
 Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA WATERSTON R.;
 Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U11029; AAA19242.1;
 DR PFAM; PF00069; Pkinase; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 SQ SEQUENCE 357 AA; 40702 MW; 16C59B90 CRC32;

Query Match 72.7%; Score 40; DB 5; Length 357;

Best Local Similarity 87.5%; Pred. No. 8.7;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTV 9

|||||

Db 301 WVTADGTV 308

RESULT 2

Q19522
ID Q19522 PRELIMINARY; PRT; 2761 AA.
AC Q19522; Q20718;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE F53B7.5 PROTEIN.
GN F53B7.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150719.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; 272510; CAA96634.1; -;
DR EMBL; 272507; CAA96634.1; JOINED.
DR EMBL; 272510; CAA96634.1; -;
DR EMBL; 272510; CAA96634.1; JOINED.
DR PFAM; PF00090; tsp_1; 1.
SQ SEQUENCE 2761 AA; 282475 MW; AF4F8A19 CRC32;

Query Match 72.7%; Score 40; DB 5; Length 2761;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTV 9
||| ||||
Db 2638 WVTBDGTV 2645

RESULT 3
Q52608
ID Q52608 PRELIMINARY; PRT; 230 AA.
AC Q52608;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)
DE HYPOTHETICAL 25.4 KD PROTEIN.
GN VIRB.
OS Agrobacterium tumefaciens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87041440.
RA WINANS S.C., EBERT P.R., STACHEL S.E., GORDON M.P., NESTER E.W.;
RT "A gene essential for Agrobacterium virulence is homologous to a
family of positive regulatory loci.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:8278-8282(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88186901.
RA WARD J.E., AKIYOSHI D.E., REGIER D., DATTA A., GORDON M.P.,
RA NESTER E.W.;

RT "Characterization of the virB operon from an Agrobacterium
RT tumefaciens Ti plasmid.";
RL J. Biol. Chem. 263:5804-5814(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90170994.
RA WARD J.E., AKIYOSHI D.E., REGIER D., DATTA A., GORDON M.P.,
RA NESTER E.W.;
RT "Correction: characterization of the virB operon from Agrobacterium
RT tumefaciens Ti plasmid.";
RL J. Biol. Chem. 264:4768-4768(1990).
DR EMBL; J03216; AAA88653.1; -;
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 230 AA; 25396 MW; 690292E4 CRC32;

Query Match 70.9%; Score 39; DB 2; Length 230;
Best Local Similarity 66.7%; Pred. No. 8.2;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
||| |||||
Db 66 WIRADGTV 74

RESULT 4
P93185
ID P93185 PRELIMINARY; PRT; 427 AA.
AC P93185;
DT 01-MAY-1997 (TRENBLrel. 03, Created)
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE (1,4)-BETA-XYLAN ENDOHYDROLASE, ISOENZYME X-I (EC 3.2.1.8)
DE (ENDO-1,4-BETA-XYLANASE) (1,4-BETA-D-XYLAN XYLANOXYDROLASE).
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euryphytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Hordeum.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-CV. HIMALAYA;
RX MEDLINE; 97071686.
RA BANIK M., GARRETT T.P., FINCHER G.B.;
RT "Molecular cloning of cDNAs encoding (1-->4)-beta-xylan
RT endohydrolases from the aleurone layer of germinated barley (Hordeum
RT vulgare).";
RL Plant Mol. Biol. 31:1163-1172(1996).
CC -|- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
CC LINKAGES IN XYLANS.
DR EMBL; U59312; AAB38389.1; -;
DR MENDEL; 8611; HORVU; 1384.1.
DR PFAM; PF00331; Glyco_hydro_10; 2.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 427 AA; 48076 MW; E7655C61 CRC32;

Query Match 70.0%; Score 38.5; DB 10; Length 427;
Best Local Similarity 81.8%; Pred. No. 20;
Matches 9; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AW-VTADGTVN 10
||| |||||
Db 348 AWLVADGTVN 358

RESULT 5
P93186
ID P93186 PRELIMINARY; PRT; 377 AA.
AC P93186;
DT 01-MAY-1997 (TRENBLrel. 03, Created)
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE (1,4)-BETA-XYLAN ENDOHYDROLASE, ISOENZYME X-II (EC 3.2.1.8)

DE (ENDO-1,4-BETA-XYLANASE) (1,4-BETA-D-XYLAN XYLANOXYDROLASE)
 DE (FRAGMENT).
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Hordeum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. HIMALAYA;
 RX MEDLINE; 97071686.
 RA BANIK M., GARRETT T.P., FINCHER G.B.;
 RT "Molecular cloning of cDNAs encoding (1->4)-beta-xylan
 RT endohydrolases from the aleurone layer of germinated barley (Hordeum
 RT vulgare).";
 RL Plant Mol. Biol. 31:1163-1172(1996).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
 CC LINKAGES IN XYLANS.
 DR ENBL; U59313; AAB38390.1; -.
 DR MENDEL; 8612; HORVU:1384;2.
 DR PFAM; PF00331; Glyco_hydro_10; 2.
 KW Hydrolase; Glycosidase.
 FT NON_TER 377 377
 SQ SEQUENCE 377 AA; 42223 MW; C9E1D18C CRC32;

Query Match 70.0%; Score 38.5; DB 10; Length 377;
 Best Local Similarity 81.8%; Pred. No. 17;
 Matches 9; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AW-VTADGTVN 10
 ||| |
 DB 343 AWLVADGTVN 353

RESULT 6.
 P93187 PRELIMINARY; PRT; 427 AA.
 ID P93187
 AC P93187;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE XLAN ENDOHYDROLASE ISOENZYME X-I (EC 3.2.1.6)
 DE (ENDO-1,4-BETA-XYLANASE) (1,4-BETA-D-XYLAN XYLANOXYDROLASE).
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Hordeum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 97218032.
 RA BANIK M., LI C.D., LANGRIDGE P., FINCHER G.B.;
 RT "Structure, hormonal regulation, and chromosomal location of genes
 RT encoding barley (1->4)-beta-xylan endohydrolases.";
 RL Mol. Gen. Genet. 233:599-608(1997).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
 CC LINKAGES IN XYLANS.
 DR ENBL; U73749; AAB51668.1; -.
 DR MENDEL; 8613; HORVU:1384;3.
 DR PFAM; PF00331; Glyco_hydro_10; 2.
 KW Hydrolase; Glycosidase.
 SQ SEQUENCE 427 AA; 47961 MW; 31C058C0 CRC32;

Query Match 70.0%; Score 38.5; DB 10; Length 427;
 Best Local Similarity 81.8%; Pred. No. 20;
 Matches 9; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AW-VTADGTVN 10
 ||| |
 DB 348 AWLVADGTVN 358

RESULT 7

Q54823 PRELIMINARY; PRT; 438 AA.
 ID Q54823;
 AC Q54823;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE DNRP.
 DE DNRQ.
 GN DNRQ.
 OS Streptomyces peucetius.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96062261.
 RA OTTEN S.L., LIU X., FERGUSON J., HUTCHINSON C.R.;
 RT "Cloning and characterization of the Streptomyces peucetius dnrQS
 RT genes encoding a daunosamine biosynthesis enzyme and a glycosyl
 RT transferase involved in daunorubicin biosynthesis.";
 RL J. Bacteriol. 177:6688-6692(1995).
 DR ENBL; L47164; AAD15266.1; -.
 DR PFAM; PF00067; p450; 1.
 SQ SEQUENCE 438 AA; 46316 MW; 5B0BD8C8 CRC32;

Query Match 69.1%; Score 38; DB 2; Length 438;
 Best Local Similarity 66.7%; Pred. No. 25;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
 ||| |
 DB 69 WVTADGGIS 77

RESULT 8
 P95198 PRELIMINARY; PRT; 140 AA.
 ID P95198;
 AC P95198;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE HYPOTHETICAL 15.4 KD PROTEIN.
 GN MTCY04D9.02.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA GENTILES S., CHURCHER C.M.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA BARRELL B.G., RAJANDREAM M.A.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE; 96181548.
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RT "An integrated map of the genome of the tubercle bacillus,
 RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
 RT leprae.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
 DR ENBL; Z84725; CAB06596.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 140 AA; 15392 MW; A7055A4E CRC32;

Query Match 67.3%; Score 37; DB 2; Length 140;
 Best Local Similarity 66.7%; Pred. No. 11;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 WVTADGTVN 10

Db 53 WATSDGTHN 61

RESULT 9

ID O29358 PRELIMINARY; PRT; 403 AA.
AC O29358;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE SUCCINYL-DIAMINOPIMELATE DESUCCINYLASE (DAPE-2).
GN AF0904
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE; 98049343
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBEK R., GOCCAYNE J.D., WEIDMAN J.F., McDONALD L., UTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTACH P., KATNE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESSE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic,
RT sulphate-reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL; AE001041; AAB90336.1; -.
DR TIGR; AF0904; -.
KW Hypothetical protein.
SQ SEQUENCE 403 AA; 45518 MW; 9F531A8F CRC32;

Query Match 65.5%; Score 36; DB 1; Length 403;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AWVTADGTVN 10

Db 367 AWCTADGVAH 376

RESULT 10

ID Q55218 PRELIMINARY; PRT; 99 AA.
AC Q55218;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE ACYLTRANSFERASE (DAUA) (FRAGMENT).
GN DAUQ.
OS Streptomyces sp.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C5;
RX MEDLINE; 95138010.
RA DICKENS M.L., YE J., STROHL W.R.;
RT "Analysis of clustered genes encoding both early and late steps in
RT daunomycin biosynthesis by Streptomyces sp. strain C5."
RL J. Bacteriol. 177:536-543(1995).
DR EMBL; L35154; AAB16940.1; -.

FT NON_TER 99 99
SQ SEQUENCE 99 AA; 10533 MW; 39A8C5B4 CRC32;

Query Match 65.5%; Score 36; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 WVTADG 7

Db 71 WVTADG 76

RESULT 11

O33560 PRELIMINARY; PRT; 333 AA.
ID O33560;
AC O33560;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 38.5 KD PROTEIN
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WS8;
RX MEDLINE; 98086113.
RA HAMBLIN P.A., MAGUIRE B.A., GRISHANIN R.N., ARMITAGE J.P.;
RT "Evidence for two chemosensory pathways in Rhodobacter sphaeroides."
RL Mol. Microbiol. 26:1083-1096(1997).
DR EMBL; AJ000977; CAA04434.1; -.
KW Hypothetical protein.
SQ SEQUENCE 333 AA; 38457 MW; A8C52A2C CRC32;

Query Match 65.5%; Score 36; DB 2; Length 333;
Best Local Similarity 71.4%; Pred. No. 43;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 WVTADGT 8

Db 33 WITADGS 39

RESULT 12

O69822 PRELIMINARY; PRT; 2314 AA.
ID O69822;
AC O69822;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE PUTATIVE SECRETED PROTEIN.
GN SCIA6.17C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA OLIVER K., HARRIS D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL MOL. Microbiol. 21:77-96(1996).
 DR EMBL; AL023496; CAA18915.1; -.
 DR PFAM; PF00041; fn3; 2.
 DR PROSITE; PS01157; ACID_PHOSPH_CLA; 1.
 SQ SEQUENCE 2314 AA; 242147 MW; 69E20F21 CRC32;

Query Match 65.5%; Score 36; DB 2; Length 2314;
 Best Local Similarity 75.0%; Pred. No. 3.6e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AWVTADGT 8
 II IIII
 DB 543 AWTADGT 550

RESULT 13
 F74440 PRELIMINARY; PRT; 4199 AA.
 AC P74440;
 DT 01-FEB-1997 (TReMBLrel. 02, Created)
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
 DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
 DE HYPOTHETICAL 442.4 KD PROTEIN.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC6803;
 RA TABATA S.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PC6803;
 RX MEDLINE; 97061201.
 RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
 RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
 RA TABATA S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D90915; BAA18541.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 4199 AA; 442438 MW; C6EE9EB8 CRC32;

Query Match 65.5%; Score 36; DB 2; Length 4199;
 Best Local Similarity 55.6%; Pred. No. 7e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 WVTADGTVN 10
 I: IIII:I
 DB 2251 WLAADGSLN 2259

RESULT 14
 Q62831 PRELIMINARY; PRT; 544 AA.
 AC Q62831;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-JAN-1999 (TReMBLrel. 09, Last sequence update)
 DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)
 DE NEUROGLYCAN C PRECURSOR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
 RX MEDLINE; 96070781.
 RA WATANABE E., MAEDA N., MATSUI F., KUSHIMA Y., NODA M., OOHIRA A.;
 RT "Neuroglycan C, a novel membrane-spanning chondroitin sulfate
 proteoglycan that is restricted to the brain.";
 RL J. Biol. Chem. 270:26876-26882(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
 RA WATANABE E.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
 RA TOKITA Y., OOHIRA A.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
 RA TOKITA Y., OOHIRA A.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U33553; AAC98537.1; -.
 KW Signal.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 544 NEUROGLYCAN C.
 SQ SEQUENCE 544 AA; 58083 MW; 1363F7BA CRC32;

Query Match 65.5%; Score 36; DB 1; Length 544;
 Best Local Similarity 75.0%; Pred. No. 74;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AWVTADGT 8
 II: IIII
 DB 105 AWLEADGT 112

RESULT 15
 O41347 PRELIMINARY; PRT; 725 AA.
 ID O41347;
 AC O41347;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
 DE 82 KDA PROTEIN.
 OS Tobacco necrosis virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Necrovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D;
 RX MEDLINE; 97335247.
 RA MOLNAR A., HAVELDA Z., DALMAY T., SZUTORISZ H., BURGANY J.;
 RT "Complete nucleotide sequence of tobacco necrosis virus strain DH and
 genes required for RNA replication and virus movement.";
 RL J. Gen. Virol. 78:1235-1239(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D;
 RA MOLNAR A.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U62546; AAC57946.1; -.
 SQ SEQUENCE 725 AA; 82384 MW; 92C66B07 CRC32;

Query Match 65.5%; Score 36; DB 12; Length 725;
 Best Local Similarity 87.5%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AWVTADGT 8
 IIIIIII
 DB 613 AWVTAVGT 620

Search completed: November 13, 1999, 12:55:31
Job time: 3030 sec

Result No.	Query	Score	Match	Length	DB	ID	Description
1	115	100.0	21	1	W01032	Mycoplasma 52-54 k	
2	49	42.6	782	1	R15625	Cellulase AE-1. Ce	
3	45.5	39.6	847	1	P91071	N-alpha-acetyl tra	
4	45.5	39.6	854	1	R12042	N-alpha-acetyltran	
5	44	38.3	889	1	W68208	M. catarrhalis str	
6	43	37.4	474	1	Y00180	Enterococcus faeca	
7	43	37.4	431	1	Y00181	Enterococcus faeca	
8	42	36.5	579	1	R29079	Porcine Parvovirus	
9	42	36.5	196	1	R39894	Antigen from clust	
10	42	36.5	662	1	W98677	H. pylori GHPO 564	
11	41	35.7	208	1	R51896	Chitin binding pro	
12	41	35.7	211	1	R51897	Chitin binding pro	
13	41	35.7	10	1	W01031	Mycoplasma 52-54 k	
14	41	35.7	916	1	W85612	Secreted protein c	
15	40	34.8	1035	1	R04025	Pol gene product o	
16	40	34.8	512	1	R37347	Truncated secreted	
17	40	34.8	493	1	R95562	Serine threonine k	
18	40	34.8	279	1	W00768	Thermotaxis G206N	
19	40	34.8	502	1	W58861	T. halophilus xylu	
20	40	34.8	75	1	W27905	Staphylococcus aur	
21	40	34.8	747	1	W60487	Human TRIDENT tran	
22	40	34.8	469	1	W5211	MMP19 protein sequ	
23	39.5	34.3	283	1	P81166	D-amino acid trans	
24	39.5	34.3	706	1	R75647	Thermophilic bacte	
25	39	33.9	448	1	P93343	Gene encoding the	
26	39	33.9	373	1	R05751	Lactic oxidase (LO	
27	39	33.9	503	1	R14768	Metastasis-specifi	
28	39	33.9	453	1	R40856	43kd regression as	
29	39	33.9	262	1	R44002	Acidic protease is	
30	39	33.9	482	1	R44006	Acidic protease V2	
31	39	33.9	1144	1	R76059	Mycoplasma pirum a	
32	39	33.9	19	1	R78756	Mycoplasma pirum a	
33	39	33.9	432	1	R67582	Cancer metastasis	
34	39	33.9	374	1	W14476	Lactate oxidase va	
35	39	33.9	374	1	W14477	Lactate oxidase va	
36	39	33.9	374	1	W14478	Lactate oxidase va	
37	39	33.9	538	1	W10026	Partial sporozoite	
38	39	33.9	763	1	W60591	Human hepatocyte n	
39	39	33.9	292	1	Y00102	Enterococcus faeca	
40	39	33.9	270	1	Y00103	Enterococcus faeca	
41	38.5	33.5	439	1	W97693	Staphylococcus aur	
42	38	33.0	1114	1	R21999	M17 antigen encode	
43	38	33.0	1529	1	R41732	High molecular wei	

Cellulase AE-1 for e.g. mfr. of pharmaceuticals and foodstuffs -
 Claim 2; of opt. pH when carboxymethylcellulose is used as substrate.
 Claim 2; Fig 3; 8pp; Japanese.
 The sequence was deduced from the gene which was sequenced from
 plasmid, PAEC 1, prepd. by ligating chromosomal DNA confg. the
 gene (obtd. from Acromonas) into pUC18. The protein has amol. wt.
 of 81,000 (SDS-PAGE) and an optimum pH near to 5 when carboxy-
 methylcellulose is the substrate. The N-terminal sequence: GHADT
 has been confirmed by Edman degradation. The gene can be used to
 produce recombinant enzyme which is used for the effective utilis-

RESULT 7

SQ Sequence 579 AA;

SQ Sequence 579 AA;

Query Match 36.5%; Score 42; DB 1; Length 579;
 Best Local Similarity 44.4%; Pred. No. 1.2e+02;
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 2 IVTADGTVDNKNQWVR 19
 : ||| ||::|| :
 Db 357 VPTADQYNDDEPNGAIR 374

RESULT 9

W89894 standard; Protein; 196 AA.
 AC W89894;
 DT 18-FEB-1999 (first entry)
 DE Antigen from cluster 27
 KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;
 KW Peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
 OS Helicobacter pylori.
 PN W09849314-A2.
 PD 05-NOV-1998.
 PF 27-APR-1998; U08487.
 PR 14-OCT-1997; US-061958.
 PR 25-APR-1997; US-045107.
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Chow TP, Fry KE, Lim MY, McAttee CP;
 DR WPI; 99-009433/01.
 PT New Helicobacter pylori antigens and related nucleic acid sequences
 PT - useful in serological diagnosis and protective vaccines, providing
 PT long-lasting immune response
 PS Claim 1; Page 235; 402pp; English.
 CC The present sequence represents a Helicobacter pylori antigenic protein
 CC that is characterised by immunoreactivity with H. pylori-positive
 CC antisera. The proteins are highly immunogenic and induce a long-lasting
 CC immune response that persists even after antimicrobial treatment. In
 CC antibody detection assays, on sera, plasma, urine, saliva etc., they are
 CC highly sensitive and specific. The specification also describes 69
 CC previously unrecognised immunogenic cluster families. H. pylori antigens
 CC are used to detect H. pylori-specific antibodies, for diagnosing
 CC infection or to confirm eradication of infection, and in vaccines to
 CC protect against H. pylori infection and related diseases (gastritis,
 CC peptic ulcer, gastric adenocarcinoma/lymphoma).
 SQ Sequence 196 AA;

Query Match 36.5%; Score 42; DB 1; Length 196;
 Best Local Similarity 53.8%; Pred. No. 37;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 IVTADGTVDNKNK 14
 : ||||| :
 Db 46 VITADGTFKRGKP 58

RESULT 10

W98677 standard; Protein; 662 AA.
 AC W98677;
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHP0.564 protein.
 KW GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.
 OS Helicobacter pylori.
 PN W09843478-A1.
 PD 08-OCT-1998.
 PF 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881227.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (INMR-) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
 DR WPI; 98-542293/46.
 DR N-PSDB; X14396.

PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 8; Page 1400-1403; 2054pp; English.
 CC This sequence represents a Helicobacter pylori GHP0 protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SQ Sequence 662 AA;

Query Match 36.5%; Score 42; DB 1; Length 662;
 Best Local Similarity 53.8%; Pred. No. 1.3e+02;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 IVTADGTVDNKNK 14
 : ||||| :
 Db 189 VITADGTFKRGKP 201

RESULT 11

R51896 standard; Protein; 208 AA.
 AC R51896;
 DT 09-SEP-1994 (first entry)
 DE Chitin binding protein CBP4.4.
 KW Chitin binding protein; CBP; antifungal; fungicide;
 KW beta-1,3-glucanase; transgenic plant; disease-resistance;
 KW crop improvement; tobacco; primer; polymerase chain reaction;
 KW PCR.
 OS Nicotiana tabacum.
 PN W09408009-A.
 PD 14-APR-1994.
 PF 05-OCT-1993; E02790.
 PR 05-OCT-1992; EP-203071.
 PR 13-MAY-1993; EP-201370.
 PA (MOGE-) MOGEN INT NV.
 PI Apotheker M, Bres-vloemans AA, Cornelissen BJC, Melchers LS;
 PI Ponstein AS, Sela-buurlage MB;
 DR WPI; 94-135576/16.
 DR P-PSDB; Q62455.
 PT New antifungal chitin binding protein from plants - without
 PT significant chitinase activity, showing synergistic activity with
 PT 1,3-beta-glucanase, also its nucleic acid, vectors, transformed
 PT plants, etc.
 PS Disclosure; Page 35-36; 54pp; English.
 CC cDNAs encoding chitin binding proteins (CBP) were isolated from
 CC a tobacco cv. Samsun NN library in phage lambda ZAP. Clone CBP4.4
 CC cDNA (Q62455) encoded a CBP of sequence R51896. Clone CBP5.2 cDNA
 CC (Q62456) encoded a CBP (R51897) that showed 97% identity with the
 CC CBP4.4 protein. A BamHI recognition site and an adenine-thymidine
 CC dinucleotide were introduced in front of clone CBP4.4 cDNA,
 CC creating a translation initiation codon; a BamHI recognition site
 CC was also introduced behind the gene. Primers Q62457-58 were used
 CC for these PCR-mediated processes. The sequences of the BamHI
 CC linkers are given in Q62459. CBP can be expressed in transgenic
 CC plants to improve resistance to fungal pathogens.
 SQ Sequence 208 AA;

Query Match 35.7%; Score 41; DB 1; Length 208;
 Best Local Similarity 57.1%; Pred. No. 55;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 8 TVNDNKNQWVRKY 21
 : ||| :
 Db 105 TWGNKKPLAWRKY 118

RESULT 12

```

CCC also identified (see also W01024-29 and W01033-37). Protective
CCC antigens and antibodies can be used in vaccines for preventing or
CCC treating mycoplasma infections, partic. M. hyopneumoniae
CCC infections in swine. They can also be used for diagnosis.
SQ Sequence 10 AA;

Query Match 35.7%; Score 41; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AIWTDGTVN 10
| | | | |
DDb 1 AWVTADGTVN 10

RESULT 14
W85612
IID W85612 standard; Protein; 916 AA.
AC W85612;
DE 02-MAR-1999 (first entry)
DE Secreted protein clone fh123.5.
KW Clone; secreted protein; protein factor; cytokine; lymphokine;
KW interferon; colony stimulating factor; CSF; interleukin; cloning;
KW tumour invasion; tumour suppression; immune boosting.
KW Homo sapiens.
OS Key Location/Qualifiers
FT 694..706
FT Peptide /note= "Predicted signal peptide/transmembrane
FT domain"
FT W09849302-AL.
PN 05-NOV-1998.
PD 24-APR-1998: U08336.
PF 23-APR-1998: US-065125.
PR 25-APR-1997: US-845236.
PR (GENY) GENETICS INST INC.
PR Agostino WJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PR Racie LA, Spaulding V, Treacy M;
PR WPI: 99-024059/02..
DR N-PSDB: V83137.
DR New polynucleotides encoding secreted human proteins - are derived
DR from human foetal brain, adult brain, adult blood or placenta cDNA
DR libraries, useful, e.g. as potential immunomodulators
DR Claim 24; Page 75-78; 104pp: English.
PS The nucleotide sequence (NS) of the full-length protein-coding
PS sequence of clones ci25.4 (V83132), ds228.6 (V83133), du410.5
PS (V83134), eh80_1 (V83135), er369_1 (V83136), fh123.5 (V83137),
PS im60_1 (V83138) or fr473_2 (V83139), (all clones are deposited as
PS ATCC 98415) and the proteins they encode are predicted to have
PS biological activities which would make them suitable for treating,
PS preventing or ameliorating medical conditions in humans and animals
PS for example, tumour suppression/invasion activity, immune system
PS boosting activity. The polynucleotides are also believed to be useful
PS for gene therapy.
SQ Sequence 916 AA;

Query Match 35.7%; Score 41; DB 1; Length 916;
Best Local Similarity 40.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 6 DGTVDNKNQWVRK 20
| | | | |
DDb 90 DGMVEQDKRDEWVKR 104

RESULT 15
RR04025
AC R04025 standard; protein; 1035 AA.
ID R04025;
DE 29-MAY-1989 (first entry)
DE Pol gene product of cDNA to HIV-2 RNA.
DE HIV; AIDS; Vaccine; pUC-HIV-2(GH-1).
KW

```

OS HIV-2.
 PN J01289486-A.
 PD 21-NOV-1989.
 PF 16-MAY-1988; 119024.
 PR 16-MAY-1988; JP-119024.
 PA Toa Nenryo Kogyo KK, Fuji Kebio KK.
 DR WPI; 90-005177/01.
 DR P-PSDB; R04024-30.
 PT DNA indicating complement to RNA gene -
 PT of Human Immunodeficiency Virus type 2 used for new vaccine or
 PT diagnostic for AIDS virus.
 PS Claim 2; Fig.4; 12pp; Japanese.
 CC CDNA to novel HIV-2 (GH-1) has been integrated into plasmid
 CC pUC HIV-2(GH-1). Useful for diagnosis and vaccination against the virus.
 CC Described here is the Pol gene product of the cDNA.
 CC See also Q02830.
 SQ Sequence 1035 AA;

Query Match 34.8%; Score 40; DB 1; Length 1035;
 Best Local Similarity 40.0%; Pred. NO. 4.3e+02;
 Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 6 DGTVDNDKPNQWVRK 20
 : | : | : | : |
 Db 509 EATIQKNDNQWTK 523

Search completed: November 13, 1999, 18:59:55
 Job time: 134 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:31 ; Search time 64.87 seconds
(without alignments)
3.700 Million cell updates/sec

Title: US-08-913-430-9
Perfect score: 115
Sequence: 1 AIVTADGTVDNKNPQWVRKY 21

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/PCTUS9.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	34.8	493	1 US-08-341-916-2	Sequence 2, Appl
2	40	34.8	493	2 US-08-805-166-2	Sequence 2, Appl
3	40	34.8	493	3 US-08-805-169-2	Sequence 2, Appl
4	39.5	34.3	398	2 US-08-957-365-2	Sequence 2, Appl
5	39.5	34.3	398	2 US-08-853-659A-45	Sequence 45, Appl
6	39	33.9	503	1 US-07-946-497-2	Sequence 2, Appl
7	39	33.9	374	1 US-08-625-876-2	Sequence 2, Appl
8	39	33.9	374	1 US-08-625-876-4	Sequence 4, Appl
9	39	33.9	374	1 US-08-625-876-6	Sequence 6, Appl
10	39	33.9	503	2 US-08-483-322-2	Sequence 2, Appl
11	39	33.9	538	2 US-08-541-759B-2	Sequence 2, Appl
12	39	33.9	763	2 US-08-742-753-4	Sequence 4, Appl
13	39	33.9	503	2 US-08-478-882-2	Sequence 2, Appl
14	39	33.9	339	2 US-08-892-880-3	Sequence 3, Appl
15	38.5	33.5	129	3 PCT-US95-04258-10	Sequence 10, Appl
16	38	33.0	1529	2 US-08-728-470-10	Sequence 10, Appl
17	37.5	32.6	125	2 US-08-408-095-35	Sequence 35, Appl
18	37	32.2	1958	1 US-07-945-283-2	Sequence 2, Appl
19	37	32.2	363	1 US-07-946-497-7	Sequence 7, Appl
20	37	32.2	525	1 US-08-077-939-17	Sequence 17, Appl
21	37	32.2	347	1 US-08-229-781-58	Sequence 58, Appl
22	37	32.2	347	1 US-08-630-918-58	Sequence 58, Appl
23	37	32.2	1078	1 US-08-264-534-32	Sequence 32, Appl
24	37	32.2	525	1 US-08-461-599-17	Sequence 17, Appl
25	37	32.2	525	1 US-08-461-621-17	Sequence 17, Appl
26	37	32.2	797	1 US-08-453-695A-112	Sequence 112, App
27	37	32.2	525	1 US-08-465-334-17	Sequence 17, Appl
28	37	32.2	2556	2 US-08-185-432-17	Sequence 17, Appl
29	37	32.2	363	2 US-08-483-322-7	Sequence 7, Appl
30	37	32.2	846	2 US-08-356-354-2	Sequence 2, Appl
31	37	32.2	1054	2 US-08-356-354-4	Sequence 4, Appl
32	37	32.2	908	2 US-08-356-354-6	Sequence 6, Appl
33	37	32.2	1068	2 US-08-537-210A-2	Sequence 2, Appl
34	37	32.2	1078	2 US-08-083-590A-11	Sequence 11, Appl
35	37	32.2	2556	2 US-08-083-590A-20	Sequence 20, Appl
36	37	32.2	1078	2 US-08-465-500-32	Sequence 32, Appl
37	37	32.2	797	2 US-08-268-161A-112	Sequence 112, App
38	37	32.2	1078	2 US-08-346-128-32	Sequence 32, Appl
39	37	32.2	363	2 US-08-478-882-7	Sequence 7, Appl

Sequence 112, App
Sequence 38, Appl
Sequence 13, Appl
Sequence 20, Appl
Sequence 30, Appl
Sequence 112, App

ALIGNMENTS

RESULT 1

US-08-341-916-2
; Sequence 2, Application US/08341916
; Patent No. 5614609
; GENERAL INFORMATION:
; APPLICANT: Ib ez, Carlos F.
; APPLICANT: Ryd n, Mikael
; APPLICANT: J invall, Henrik
; TITLE OF INVENTION: A No. 5614609el Serine Threonine Kinase Receptor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/341,916
; FILING DATE: Herewith
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1459.0230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-341-916-2

Query Match 34.8%; Score 40; DB 1; Length 493;
Best Local Similarity 46.7%; Pred. No. 98;
Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 VTADGTVDNKNPQW 17
Db 434 VVCDKLRPNLPQW 448

RESULT 2

US-08-805-166-2
; Sequence 2, Application US/08805166
; Patent No. 5789565
; GENERAL INFORMATION:
; APPLICANT: Ib ez, Carlos F.
; APPLICANT: Ryd n, Mikael
; APPLICANT: J invall, Henrik
; TITLE OF INVENTION: A No. 5789565el Serine Threonine Kinase Receptor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,166
FILING DATE: 24-FEB-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/341,916
FILING DATE: 15-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Goldstein, Jorge A.
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 1459.0230001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-805-166-2

Query Match 34.8%; Score 40; DB 2; Length 493;
Best Local Similarity 46.7%; Pred. No. 98;
Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Oy 3 VTADGTVDNKNPQW 17
| | | | |
Db 434 VVCDQKLRPNLPQW 448

RESULT 3
US-08-805-169-2
; Sequence 2, Application US/088051169
; Patent No. 5811245
; GENERAL INFORMATION:
; APPLICANT: Ib ez, Carlos F.
; APPLICANT: Ryd n, Mikael
; APPLICANT: J invall, Henrik
; TITLE OF INVENTION: A No. 5811245el Serine Threonine Kinase Receptor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,169
; FILING DATE: 24-FEB-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/341,916
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:

NAME: Goldstein, Jorge A.
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 1459.0230001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-805-169-2

Query Match 34.8%; Score 40; DB 2; Length 493;
Best Local Similarity 46.7%; Pred. No. 98;
Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Oy 3 VTADGTVDNKNPQW 17
| | | | |
Db 434 VVCDQKLRPNLPQW 448

RESULT 4
US-08-957-365-2
; Sequence 2, Application US/08957365
; Patent No. 5891638
; GENERAL INFORMATION:
; APPLICANT: Ib ez, Carlos F.
; APPLICANT: Ryd n, Mikael
; APPLICANT: J invall, Henrik
; TITLE OF INVENTION: A No. 5891638el Serine Threonine Kinase Receptor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,365
; FILING DATE: 24-OCT-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/805,166
; FILING DATE: 24-FEB-1997
; APPLICATION NUMBER: 08/341,916
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1459.0230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-957-365-2

Query Match 34.8%; Score 40; DB 2; Length 493;
Best Local Similarity 46.7%; Pred. No. 98;
Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 3 VTADGTVDNKNQW 17
| | | | |
Db 434 VVCDQKLRPNLQW 448

RESULT 5
US-08-853-659A-45
; Sequence 45, Application US/08853659A
; Patent No. 5925522
; GENERAL INFORMATION:
; APPLICANT: Wong, K.K.; Saffer, J.D.
; TITLE OF INVENTION: A Salmonella Sequence, Methods Of Detection
; TITLE OF INVENTION: Of A
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Paul W. Zimmerman
; ADDRESSEE: Intellectual Property Services
; ADDRESSEE: Battelle Memorial Institute
; ADDRESSEE: PNNL P.O. Box 999
; STREET: Washington Way
; CITY: Richland
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 99352
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.50", 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (WordPerfect 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/853.659A
; FILING DATE: Unknown
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: n/a
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-853-659A-45

Query Match 34.38; Score 39.5; DB 2; Length 398;
Best Local Similarity 50.0%; Pred. No. 93;
Matches 9; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Qy 3 VTAD-----GTVDNKNPN 15
: | | | | |
Db 260 ITLDSADDSGTANDNKTN 277

RESULT 6
US-07-946-497-2
; Sequence 2, Application US/07946497
; Patent No. 5506119
; GENERAL INFORMATION:
; APPLICANT: HERRLICH, Peter
; APPLICANT: PONTA, Helmut
; APPLICANT: GUENTHER, Ursula
; APPLICANT: MATZKU, Siegfried
; APPLICANT: WENGL, Achim
; TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
; TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
; TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/946,497
FILING DATE: 19921109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-946-497-2

Query Match 33.98; Score 39; DB 1; Length 503;
Best Local Similarity 46.7%; Pred. No. 1.4e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 IVTADGTVDNKNQ 16
| | | | |
Db 445 INSGNGTVDKRPSE 459

RESULT 7
US-08-625-876-2
; Sequence 2, Application US/08625876
; Patent No. 5658471
; GENERAL INFORMATION:
; APPLICANT: MINAGAWA, Hirotaka
; APPLICANT: NAKAYAMA, No. 5656471iyuki
; APPLICANT: NAKAMOTO, Shinya
; TITLE OF INVENTION: LACTATE OXIDASE WITH AN IMPROVED THERMAL
; TITLE OF INVENTION: STABILITY AND GENE OF THE SAME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/625,876
FILING DATE: 01-APR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-95947
FILING DATE: 30-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-146186
FILING DATE: 13-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J

REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: PF-1612
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 521-2297
TELEFAX: (703) 685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-625-876-2

Query Match 33.9%; Score 39; DB 1; Length 374;
Best Local Similarity 40.0%; Pred. No. 1e+02;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 IVTADGTVNDNKNQWVRKY 21
I:|||||I:|
Db 170 ILTADSTVSGNRDRDVKNF 189

RESULT 8
US-08-625-876-4
; Sequence 4, Application US/08625876
; Patent No. 5656471
; GENERAL INFORMATION:
; APPLICANT: MINAGAWA, Hirotaka
; APPLICANT: NAKAYAMA, No. 5656471lyuk1
; APPLICANT: NAKAMOTO, Shinya
; TITLE OF INVENTION: LACTATE OXIDASE WITH AN IMPROVED THERMAL
; TITLE OF INVENTION: STABILITY AND GENE OF THE SAME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,876
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-95947
; FILING DATE: 30-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-146186
; FILING DATE: 13-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: PF-1612
; TELEPHONE: (703) 521-2297
; TELEFAX: (703) 685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-625-876-4

Query Match 33.9%; Score 39; DB 1; Length 374;
Best Local Similarity 40.0%; Pred. No. 1e+02;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 IVTADGTVNDNKNQWVRKY 21
I:|||||I:|
Db 170 ILTADSTVSGNRDRDVKNF 189

RESULT 10
US-08-483-322-2
; Sequence 2, Application US/08483322
; Patent No. 5760178

Query Match 33.9%; Score 39; DB 1; Length 374;
Best Local Similarity 40.0%; Pred. No. 1e+02;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 IVTADGTVNDNKNQWVRKY 21
I:|||||I:|
Db 170 ILTADSTVSGNRDRDVKNF 189

RESULT 9
US-08-625-876-6
; Sequence 6, Application US/08625876
; Patent No. 5656471
; GENERAL INFORMATION:
; APPLICANT: MINAGAWA, Hirotaka
; APPLICANT: NAKAYAMA, No. 5656471lyuk1
; APPLICANT: NAKAMOTO, Shinya
; TITLE OF INVENTION: LACTATE OXIDASE WITH AN IMPROVED THERMAL
; TITLE OF INVENTION: STABILITY AND GENE OF THE SAME
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,876
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-95947
; FILING DATE: 30-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-146186
; FILING DATE: 13-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: PF-1612
; TELEPHONE: (703) 521-2297
; TELEFAX: (703) 685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-625-876-6

Query Match 33.9%; Score 39; DB 1; Length 374;
Best Local Similarity 40.0%; Pred. No. 1e+02;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 IVTADGTVNDNKNQWVRKY 21
I:|||||I:|
Db 170 ILTADSTVSGNRDRDVKNF 189

RESULT 10
US-08-483-322-2
; Sequence 2, Application US/08483322
; Patent No. 5760178

GENERAL INFORMATION:
APPLICANT: HERRLICH, Peter
APPLICANT: PONTA, Helmut
APPLICANT: GUENTHER, Ursula
APPLICANT: MATZKU, Siegfried
APPLICANT: WENZL, Achim
TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA
SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE PROTEINS,
TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,322
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/946,497
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16915/145
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-322-2

Query Match 33.9%; Score 39; DB 2; Length 503;
Best Local Similarity 46.7%; Pred. No. 1.4e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 IVTADGTVDNKPQ 16
I : ||||| I :
Db 445 INSGTGVDRKPSE 459

RESULT 11
US-08-541-759B-2
Sequence 2, Application US/08541759B
Patent No. 5861160
GENERAL INFORMATION:
APPLICANT: Quick, Douglas P.
APPLICANT: Welter, Mark W.
APPLICANT: Welter, Joseph
APPLICANT: Welter, Lisa M.
TITLE OF INVENTION: ISOSPORA SUIIS VACCINE
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Akzo No. 5861160el Patent Dept.
STREET: 1300 Piccard Drive, Suite 206
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/541,759B
FILING DATE: 10-OCT-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Klesner, Sharon N.
REGISTRATION NUMBER: 36,335
REFERENCE/DOCKET NUMBER: Quickla
TELEPHONE: 301-948-7400
TELEFAX: 301-948-9751
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 538 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-541-759B-2

Query Match 33.9%; Score 39; DB 2; Length 538;
Best Local Similarity 47.6%; Pred. No. 1.5e+02;
Matches 10; Conservative 2; Mismatches 3; Indels 2;

QY 2 IVTADGTVDNKPQW-VKRY 21
I : ||||| I : |||||
Db 184 LVTDTGCV-----NEWNTRKY 199

RESULT 12
US-08-742-753-4
Sequence 4, Application US/08742753
Patent No. 5861278
GENERAL INFORMATION:
APPLICANT: WONG, Gordon G.
APPLICANT: YAO, Kwok-Ming
TITLE OF INVENTION: HNF3-delta Compositions
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,753
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5277
TELEPHONE: (617) 498-8260
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA


```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04258
; FILING DATE: 06-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 109257
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: PECHT-1 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-04258-10

```

```

Query Match      33.5%; Score 38.5; DB 3; Length 129;
Best Local Similarity 42.9%; Pred. No. 39;
Matches 9; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

```

```

QY 1 AIVTADGTVNDN---KPNQWV 18
| : ||| ||| : |||
Db 103 AEILSDGHNDNFCCQVNRWV 123

```

Search completed: November 13, 1999, 10:56:32
Job time: 1361 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:37 ; Search time 251.81 Seconds
(without alignments)
5.279 Million cell updates/sec

Title: US-08-913-430-9

Perfect score: 115

Sequence: 1 AIVTADGTVDNKNPNQWVRKY 21

Scoring table: BLOSUM62

Searched: 497124 seqs, 53304441 residues

Database : Pending_Patents_AA:*
1: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115	100.0	21	13	US-08-913-430-9
2	46	40.0	191	16	US-09-248-796-19079
3	46	40.0	191	19	US-09-248-796-19079
4	44	38.3	261	16	US-09-248-796-18455
5	44	38.3	261	19	US-09-248-796-18455
6	43	37.4	474	14	US-09-071-035-342
7	43	37.4	431	14	US-09-071-035-344
8	43	37.4	496	15	US-09-134-000-5286
9	43	37.4	202	16	US-09-270-767-34039
10	43	37.4	202	16	US-09-270-767-49256
11	43	37.4	469	17	US-09-328-352-4250
12	43	37.4	163	24	US-09-147-499-5113
13	42.5	37.0	348	23	US-09-417-507-26290
14	42	36.5	662	1	PCT-US98-06371-954
15	42	36.5	564	1	PCT-US99-16164-2
16	42	36.5	662	13	US-08-903-615A-220
17	42	36.5	196	14	US-09-065-756A-386
18	42	36.5	564	15	US-09-116-750-2
19	42	36.5	315	16	US-09-248-796-15178

20	42	36.5	386	17	US-09-328-352-7130
21	42	36.5	315	19	US-09-096-409-15178
22	41.5	36.1	234	23	US-09-417-507-28141
23	41.5	36.1	583	23	US-09-417-507-41715
24	41	35.7	208	7	US-08-411-640-8
25	41	35.7	211	7	US-08-411-640-10
26	41	35.7	556	12	US-08-869-696-2
27	41	35.7	395	12	US-08-869-696-3
28	41	35.7	348	12	US-08-869-696-4
29	41	35.7	497	12	US-08-869-696-9
30	41	35.7	804	13	US-08-909-934-2
31	41	35.7	10	13	US-08-913-430-8
32	41	35.7	208	13	US-08-935-886-8
33	41	35.7	211	13	US-08-935-886-10
34	41	35.7	916	14	US-09-065-125-12
35	41	35.7	270	16	US-09-248-796-1842
36	41	35.7	271	16	US-09-248-796-18869
37	41	35.7	70	16	US-09-248-796-22184
38	41	35.7	265	16	US-09-270-767-44092
39	41	35.7	165	16	US-09-270-767-45836
40	41	35.7	110	17	US-09-328-352-8212
41	41	35.7	270	19	US-09-096-409-1842
42	41	35.7	271	19	US-09-096-409-18869
43	41	35.7	70	19	US-09-096-409-22184
44	41	35.7	87	23	US-09-417-507-41135
45	41	35.7	110	24	US-09-160-209-4049

ALIGNMENTS

RESULT 1
US-08-913-430-9
; Sequence 9, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; EARLIER FILING DATE: 1997-12-09
; EARLIER FILING DATE: 1996-03-15
; EARLIER FILING DATE: PN 1789
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0 - beta
; SEQ ID NO 9
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-9

Query Match	100.0%	Score 115;	DB 13;	Length 21;
Best Local Similarity	100.0%	Pred. No. 1.5e-11;		
Matches	21;	Conservative	0;	Mismatches
		Indels	0;	Gaps
				0;
QY	1	AIVTADGTVDNKNPNQWVRKY	21	
Db	1	AIVTADGTVDNKNPNQWVRKY	21	
RESULT 2				
US-09-248-796-19079				
; Sequence 19079, Application US/09248796				
; GENERAL INFORMATION:				
; APPLICANT: Keith Weinstock et al				
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI				
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS				
; FILE REFERENCE: 107196.132				
; CURRENT APPLICATION NUMBER: US/09/248,796				

; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 19079
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-19079

Query Match 40.0%; Score 46; DB 16; Length 191;
Best Local Similarity 46.7%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 7 GTVNDNKPQWVRKY 21
| | | | | | | | | |
DB 112 GRVNTPENWKDY 126

RESULT 3
US-60-096-409-19079
; Sequence 19079, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096.409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 19079
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-19079

Query Match 40.0%; Score 46; DB 19; Length 191;
Best Local Similarity 46.7%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 7 GTVNDNKPQWVRKY 21
| | | | | | | | | |
DB 112 GRVNTPENWKDY 126

RESULT 4
US-09-248-796-18455
; Sequence 18455, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 18455
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (255)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796-18455

Query Match 38.3%; Score 44; DB 16; Length 261;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 TADGTVDNKPQWVR 19
| | | | | | | | | |

DB 139 TAEETIEDNEPNTTR 154

RESULT 5
US-60-096-409-18455
; Sequence 18455, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096.409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 18455
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (255)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are un
US-60-096-409-18455

Query Match 38.3%; Score 44; DB 19; Length 261;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 TADGTVDNKPQWVR 19
| | | | | | | | | |
DB 139 TAEETIEDNEPNTTR 154

RESULT 6
US-09-071-035-342
; Sequence 342, Application US/09071035
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/071.035
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 342:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-09-071-035-342

Query Match 37.4%; Score 43; DB 14; Length 474;
Best Local Similarity 41.2%; Pred. No. 1.2e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 IVTADGTVNDKNKNQWV 18

:: |||:| | ||

Db 229 VLNDGTVNKEKLTWV 245

RESULT 7

US-09-071-035-344

; Sequence 344, Application US/09071035

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage.

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/071,035

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: A. Anders Brooks

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB369P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 344:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 431 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-071-035-344

Query Match 37.4%; Score 43; DB 14; Length 431;

Best Local Similarity 41.2%; Pred. No. 1.1e+02;

Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 IVTADGTVNDKNKNQWV 18

:: |||:| | ||

Db 186 VLNDGTVNKEKLTWV 202

RESULT 8

US-09-134-000-5286

; Sequence 5286, Application US/09134000A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS

; TITLE OF INVENTION: FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-005

; CURRENT APPLICATION NUMBER: US/09/134,000A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 6810
; SEQ ID NO 5286
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000-5286

Query Match 37.4%; Score 43; DB 15; Length 496;
Best Local Similarity 41.2%; Pred. No. 1.3e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 IVTADGTVNDKNKNQWV 18

:: |||:| | ||

Db 251 VLNDGTVNKEKLTWV 267

RESULT 9

US-09-270-767-34039

; Sequence 34039, Application US/09270767

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 34039

; LENGTH: 202

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-270-767-34039

Query Match 37.4%; Score 43; DB 16; Length 202;

Best Local Similarity 53.8%; Pred. No. 46;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 ADGTVNDKNKNQW 17

:: |||:| | ||

Db 182 AKGIVNEDLPNSW 194

RESULT 10

US-09-270-767-49256

; Sequence 49256, Application US/09270767

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 49256

; LENGTH: 202

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-09-270-767-49256

Query Match 37.4%; Score 43; DB 16; Length 202;

Best Local Similarity 53.8%; Pred. No. 46;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 ADGTVNDKNKNQW 17

:: |||:| | ||

Db 182 AKGIVNEDLPNSW 194

RESULT 11
US-09-328-352-4250
; Sequence 4250, Application US/09328352
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breston et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4250
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4250

Query Match 37.4%; Score 43; DB 17; Length 469;
Best Local Similarity 53.3%; Pred. No. 1.2e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 VTADGTVDNKNPQW 17
| | | | | : | |
DB 189 VYADGTLNPKRAELW 203

RESULT 12
US-60-147-499-5113
; Sequence 5113, Application US/60147499
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.034PR2
; CURRENT APPLICATION NUMBER: US/60/147,499
; CURRENT FILING DATE: 1999-08-05
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 5113
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 148
; OTHER INFORMATION: Xaa = His,Pro
US-60-147-499-5113

Query Match 37.4%; Score 43; DB 24; Length 163;
Best Local Similarity 47.1%; Pred. No. 36;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 IVTADGTVDNKNPQW 18
| | | | | : | |
DB 140 ILNADGPAXLELPQWL 156

RESULT 13
US-09-417-507-26290
; Sequence 26290, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; FILE OF INVENTION: FUMIGATUS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 26290
; LENGTH: 348

; TYPE: PRT
; ORGANISM: A.fumigatus
US-09-417-507-26290

Query Match 37.0%; Score 42.5; DB 23; Length 348;
Best Local Similarity 57.1%; Pred. No. 1e+02;
Matches 8; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 7 GTVNDN-KPNQWVR 19
| | | | | : | | | |
DB 107 GTLDDREPQWFR 120

RESULT 14
PCT-US98-06371-954
; Sequence 954, Application PC/TUS9806371
; GENERAL INFORMATION:
; APPLICANT: MERIEUX ORAVAX SOCIETE EN NOM COLLECTIF PASTEUR
; APPLICANT: MERIEUX SERUMS ET VACCINS S.A. HUMAN GENOME
; APPLICANT: SCIENCES, INC.
; TITLE OF INVENTION: Identification of Polynucleotides
; FILE OF INVENTION: Encoding Novel Helicobacter Polypeptides in the Helicobacte
; TITLE OF INVENTION: Genome
; NUMBER OF SEQUENCES: 1376
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US98/06371
; FILING DATE: 01-APR-98
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/833,457
; FILING DATE: 01-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/881,227
; FILING DATE: 24-JUN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/902,615
; FILING DATE: 29-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06132/041W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 954:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 662 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US98-06371-954

Query Match 36.5%; Score 42; DB 1; Length 662;
Best Local Similarity 53.8%; Pred. No. 2.5e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 IVTADGTVDNKNP 14

Db 189 VITADGTFRKGP 201
:::|||||

RESULT 15
PCT-US99-16164-2
; Sequence 2, Application PC/TUS9916164
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: BANDMAN, Olga
; APPLICANT: TANG, Y. Tom
; APPLICANT: CORLEY, Neil C.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: BAUGHN, Mariah R.
; TITLE OF INVENTION: SCAD-RELATED MOLECULES
; FILE REFERENCE: PF-0559 PCT
; CURRENT APPLICATION NUMBER: PCT/US99/16164
; CURRENT FILING DATE: 1999-07-16
; EARLIER APPLICATION NUMBER: 09/116,750; Unassigned
; EARLIER FILING DATE: 1998-07-16; 1998-07-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 564
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte Clone No: 2060002
; PUBLICATION INFORMATION:
PCT-US99-16164-2

Query Match 36.5%; Score 42; DB 1; Length 564;
Best Local Similarity 47.1%; Pred. No. 2.1e+02;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 IVTADGTVNDNKNQWV 18
I: |||
Db 167 ILNADGPAPLEPNOWL 183

Search completed: November 13, 1999, 05:08:38
Job time: 10864 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:08:01 ; Search time 75.45 Seconds
(without alignments)
11.151 Million cell updates/sec

Title: US-08-913-430-9
Perfect score: 115
Sequence: 1 AIVTAGTVNDKNPNQWRKY 21

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	46.1	483	2 S41853	centromere/microtu
2	48	41.7	823	2 S44873	ZC21.2 protein - C
3	47	40.9	140	1 R3B512	ribosomal protein
4	47	40.9	138	2 C69700	ribosomal protein
5	47	40.9	139	2 S73932	ribosomal protein
6	47	40.9	139	2 F64209	ribosomal protein
7	46	40.0	1324	1 VGIH59	E2 glycoprotein pr
8	45.5	39.6	854	1 XYBYT1	protein N-acetyltr
9	45	39.1	455	2 S03813	beta-glucosidase (
10	45	39.1	316	2 S52753	hypothetical prote
11	44	38.3	137	2 S26680	ribosomal protein
12	44	38.3	117	2 S19123	hypothetical prote
13	44	38.3	334	2 S32166	ccsa protein - gol
14	43	37.4	1235	1 VGIHMT	E2 glycoprotein pr
15	43	37.4	1376	1 VGIHJ2	E2 glycoprotein pr
16	43	37.4	1376	1 JQ1534	E2 glycoprotein pr
17	43	37.4	359	2 JT0391	site-specific meth
18	43	37.4	322	2 D64996	hypothetical prote
19	43	37.4	747	2 D70802	hypothetical prote
20	43	37.4	1069	2 T00377	KIAA0642 protein -
21	42	36.5	723	1 VCPVPP	coat protein vp1 -
22	42	36.5	729	1 VCPVNA	coat protein vp1 -
23	42	36.5	729	1 A60006	hemagglutinin prec
24	42	36.5	565	1 HMIVB1	gamma-glutamyl pho
25	42	36.5	443	2 C70392	ferrochelatase (EC
26	42	36.5	466	2 A54125	acetyl-CoA synthet
27	42	36.5	582	2 E64550	capsid protein - h
28	42	36.5	539	2 S40111	yeast protein - h
29	42	36.5	244	2 S70176	probable polyketid
30	42	36.5	1402	2 D70634	probable polyketid
31	42	36.5	1293	2 T01512	probable purine nu
32	42	36.5	244	2 S37984	occlusion-derived
33	41.5	36.1	233	3 T00202	DNA primase - sate
34	41	35.7	777	1 RPBP4	valine--trna ligas
35	41	35.7	875	2 A70192	DNA primase - phag
36	41	35.7	777	2 C41830	probable enoyl-CoA
37	41	35.7	255	2 D64890	luxR protein - vib
38	41	35.7	252	2 B33538	regulatory protein
39	41	35.7	252	2 B33538	regulatory protein

ALIGNMENTS

RESULT 1

S41853
centromere/microtubule-binding protein CBF5 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L9470.11; protein YLR175W
C:Species: Saccharomyces cerevisiae
C:Date: 31-Mar-1992 #sequence_revision 14-Sep-1994 #text_change 06-Feb-1998
C:Accession: S41853; S45443; S51420
R:Jiang, W.; Middleton, K.; Yoon, H.J.; Fouquet, C.; Carbon, J.
Mol. Cell. Biol. 13, 4884-4893, 1993
A:Title: An essential yeast protein, CBF5p, binds in vitro to centromeres and microtu
A:Reference number: S41853; MUID:93330283
A:Accession: S41853
A:Molecule type: DNA
A:Residues: 1-483 <JIA>
A:Cross-references: EMBL:L12351; NID:g311096; PID:g311097
A:Accession: S45443
A:Molecule type: protein
A:Residues: 36-45;142-149;168-175;385-401 <J12>
R:Wohldmann, P.
submitted to the EMBL Data Library, November 1994
A:Description: The sequence of S. cerevisiae cosmid 9470.
A:Reference number: S51414
A:Accession: S51420
A:Molecule type: DNA
A:Residues: 1-483 <WOH>
A:Cross-references: EMBL:U17246; NID:g577192; PID:g577203; MIPS:YLR175W
C:Genetics:
A:Gene: SGD:CBF5
A:Cross-references: SGD:S0004165; MIPS:YLR175W
A:Map position: 12R
F:434-463/Region: Lys-Lys-Glu/Asp repeats

Query Match 46.1%; Score 53; DB 2; Length 483;
Best Local Similarity 53.3%; Pred. No. 1.2;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 7 GTVNDKNPNQWRKY 21

| | | | | | | | | |

Db 372 GRVNTPEQWKKEY 386

RESULT 2

S44873
ZC21.2 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 04-Sep-1998
C:Accession: S44873
R:Du, Z.; Waterston, R.
submitted to the EMBL Data Library, May 1993
A:Description: Sequence of the C. elegans cosmid ZC21.
A:Reference number: S44647
A:Accession: S44873
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-823 <DUZ>
A:Cross-references: EMBL:L16685; NID:g289729; PID:g289732
C:Genetics:
A:Introns: 47/2; 91/3; 144/1; 215/2; 344/3; 394/3; 440/3; 506/2; 566/3; 756/3
C:Superfamily: TRPC3 protein
C:Keywords: transmembrane protein

Search completed: November 13, 1999, 12:08:03
Job time: 2075 sec

Result No.	Score	Query %		Length	DB	ID	Description
		Match	Length				
1	53	46.1	483	1	CBF5_YEAST	P33322	saccharomyc
2	48	41.7	823	1	YN52_CAEEL	P34586	caenorhabdi
3	47	40.9	139	1	RS12_BACST	P09901	bacillus st
4	47	40.9	137	1	RS12_BACSU	P24732	bacillus su
5	47	40.9	139	1	RS12_MYCGE	P47373	mycoplasma
6	47	40.9	139	1	RS12_MYCPN	P75546	mycoplasma
7	47	40.9	137	1	RS12_STPAU	P48942	staphylococ
8	46	40.0	479	1	CBP5_CANAL	O43101	candida alb
9	46	40.0	1324	1	VG12_CVMA5	P11224	murine coro
10	45.5	39.6	250	1	LUX5_VIBF1	P35327	vibrio fisc
11	45.5	39.6	853	1	NAT1_YEAST	P12945	saccharomyc
12	45	39.1	455	1	BGL5_CALSA	P10482	caldocellum
13	44	38.3	334	1	CHL1_OILU	Q32742	olisthodisc
14	44	38.3	137	1	RS12_STRPN	P30891	streptococc
15	43	37.4	359	1	MTH1_HAEIN	P20590	haemophilus
16	43	37.4	1376	1	VG12_CVM4	P24332	murine coro
17	43	37.4	1376	1	VG12_CVM3C	Q02385	murine coro
18	43	37.4	1235	1	VG12_CVMJH	P11225	murine coro
19	43	37.4	322	1	YFBE_ECOLI	P77757	escherichia
20	43	37.4	563	1	WY03_CAEEL	Q10906	caenorhabdi
21	42	36.5	729	1	COA1_PAVP9	P33484	porcine par
22	42	36.5	729	1	COA1_PAVPK	P52501	porcine par
23	42	36.5	729	1	COA1_PAVPN	P18546	porcine par
24	42	36.5	749	1	COA2_PAVPN	P23964	porcine par
25	42	36.5	565	1	HEM1_IARUR	P17002	influenza a
26	42	36.5	466	1	HM21_ARATH	P41203	arabidopsis
27	42	36.5	445	1	NO50_CAEEL	O17919	caenorhabdi
28	42	36.5	244	1	SRPB_YEAST	P36057	saccharomyc
29	42	36.5	244	1	YENR_YEREN	P54295	yersinia en
30	41	35.7	474	1	CBF5_KULU4	O13473	kluyveromyc
31	41	35.7	539	1	COAT_LORDV	P54635	lordsdale v
32	41	35.7	435	1	FMRA_ANTEL	P10419	anthopleura
33	41	35.7	87	1	GBG_LOLFO	Q01821	loligo forbi
34	41	35.7	250	1	LUXE_VIBF1	P12746	vibrio fisc
35	41	35.7	508	1	NO60_DROME	O44081	drosophila
36	41	35.7	255	1	PAAF_ECOLI	P76082	escherichia
37	41	35.7	777	1	PRIN_BP44	P10277	bacterioph
38	41	35.7	875	1	SYV_BORBU	O51680	borrella bu
39	40.5	35.2	642	1	YNJ5_YEAST	P53932	saccharomyc
40	40	34.8	1274	1	BXF_CLOBO	P30996	clostridium
41	40	34.8	1035	1	CC66_YEAST	P32558	saccharomyc
42	40	34.8	393	1	CRTC_SCHMA	Q06814	schistosoma
43	40	34.8	214	1	KAD_MYCGE	P47417	mycoplasma

Query Match

46.18; Score 53; DB 1; Length 483;

Best Local Similarity 53.3%; Pred. No. 0.76;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 GTVNDKPNQWVRKY 21
| | | | | | | | | |
Db 372 GRVNETPEQWKEY 386

RESULT 2
YN52_CAEL
ID YN52_CAEL STANDARD; PRT; 823 AA.
AC P34586;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 95.6 KD PROTEIN ZC21.2 IN CHROMOSOME III.
GN ZC21.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS J., HILLIER L., JIER M.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RC CHANNEL (BY SIMILARITY).
CC -1- FUNCTION: COULD MEDIATE CALCIUM ENTRY AND FORM A CALCIUM PERMEANT
CC CHANNEL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE TRP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L16685; G289732; -.
DR PIR; S44873; S44873.
DR WORMPEP; ZC21.2; CEl4992.
DR PFAM; PF00023; ank; 2.
KW HYPOTHETICAL PROTEIN; IONIC CHANNEL; TRANSMEMBRANE; ION TRANSPORT;
KW CALCIUM CHANNEL.
FT TRANSMEM 375 POTENTIAL.
FT TRANSMEM 411 POTENTIAL.
FT TRANSMEM 452 POTENTIAL.
FT TRANSMEM 493 POTENTIAL.
FT TRANSMEM 536 POTENTIAL.
FT TRANSMEM 579 POTENTIAL.
FT TRANSMEM 619 POTENTIAL.
FT TRANSMEM 660 POTENTIAL.
FT TRANSMEM 714 POTENTIAL.
SQ SEQUENCE 823 AA; 95620 MW; A80CA18E CRC32;

Query Match 41.7%; Score 48; DB 1; Length 823;
Best Local Similarity 45.0%; Pred. No. 8.1;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 AIVTADGTVDNKNQWVRK 20
| | | | | | | | | |
Db 309 AILNKGNVNDNDYWASK 328

RESULT 3
RS12_BACST
ID RS12_BACST STANDARD; PRT; 139 AA.
AC P09901;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE 30S RIBOSOMAL PROTEIN S12 (BS12).
GN RPSL.
OS BACILLUS STEAROTHERMOPHILUS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91248486.
RA KIMURA M.;
RT "The nucleotide sequences of Bacillus stearothermophilus ribosomal
RT protein S12 and S7 genes: comparison with the str operon of
RT Escherichia coli.";
RL AGRIC. BIOL. CHEM. 55:207-213(1991).
RN [2]
RP SEQUENCE OF 1-138, AND SEQUENCE OF 1-78 FROM N.A.
RX MEDLINE; 87105937.
RA KIMURA M., KIMURA J.;
RT "The complete amino acid sequence of ribosomal protein S12 from
RT Bacillus stearothermophilus.";
RL FEBS LETT. 210:91-96(1987).
CC -1- FUNCTION: PROTEIN S12 IS INVOLVED IN THE TRANSLATION INITIATION
CC STEP.
CC -1- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
DR PIR; JG0007; R3BS12.
DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
DR PFAM; PF00164; S12; 1.
KW RIBOSOMAL PROTEIN.
FT INIT_MET 0
FT CONFLICT 103 103 R -> P (IN REF. 2).
FT CONFLICT 113 113 G -> A (IN REF. 2).
SQ SEQUENCE 139 AA; 15379 MW; 0786CA21 CRC32;

Query Match 40.9%; Score 47; DB 1; Length 139;
Best Local Similarity 45.0%; Pred. No. 1.7;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 IVTADGTVDNKNQWVRKY 21
| | | | | | | | | |
Db 45 VCTRVGTMTPKPKNSALRKY 64

RESULT 4
RS12_BACSU
ID RS12_BACSU STANDARD; PRT; 137 AA.
AC P21472;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE 30S RIBOSOMAL PROTEIN S12 (BS12).
GN RPSL OR STRA OR FUN.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-168;
RX MEDLINE; 97124188.
RA YASUMOTO K., LIU H., JEONG S.M., OHASHI Y., KAKINUMA S.,
RA TANAKA K., KANAMURA F., YOSHIKAWA H., TAKAHASHI H.;
RT "sequence analysis of a 50 kb region between spo0H and rrnH on the

```

RT Bacillus subtilis chromosome."
RL MICROBIOLOGY 142:3039-3046(1996).
RN [2]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN-168 / MARRU;
RX MEDLINE; 95386470.
RA BOOR K.J., DUNCAN M.L., PRICE C.W.;
RT "Genetic and transcriptional organization of the region encoding the
RL beta subunit of Bacillus subtilis RNA polymerase."
RT J. BIOL. CHEM. 270:20329-20336(1995).
RN [3]
RP SEQUENCE OF 1-20.
RX MEDLINE; 82219212.
RA HIGO K.I., OTAKA E., OSAWA S.;
RT "Purification and characterization of 30S ribosomal proteins from
RL Bacillus subtilis: correlation to Escherichia coli 30S proteins."
RL MOL. GEN. GENET. 185:239-244(1982).
CC -1- FUNCTION: PROTEIN S12 IS INVOLVED IN THE TRANSLATION INITIATION
CC STEP.
CC -1- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64127; G1644221; -.
DR EMBL; L43593; G950305; -.
DR EMBL; Z99104; E182043; -.
DR PIR; S11362; S11362.
DR SUBTILIST; BG19009; RPSL.
DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
DR PFAM; PF00164; S12; 1.
RW RIBOSOMAL PROTEIN.
FT INIT_MET 0
SQ SEQUENCE 137 AA; 15193 MW; C53EBF32 CRC32;
Query Match 40.98; Score 47; DB 1; Length 137;
Best Local Similarity 45.0%; Pred. No. 1.7;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 2 IVTDGTVNDKNKPNQWVKY 21
: | ||: ||| :|||
Db 45 VCTRVGTMTPKPNSALRKY 64
RESULT 5
ID RS12_MYCGE STANDARD; PRT; 139 AA.
AC P47333;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE 30S RIBOSOMAL PROTEIN S12.
GN RPSL OR RPS12 OR MG087.
OS MYCOPLASMA GENITALIUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE; 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."

```

```

RL SCIENCE 270:397-403(1995).
RN [2]
RP SEQUENCE OF 117-139 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE; 94075230.
RA PETERSON S.N., HU P.-C., BOTT K.F., HUTCHISON C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RL J. BACTERIOL. 175:7918-7930(1993).
CC -1- FUNCTION: PROTEIN S12 IS INVOLVED IN THE TRANSLATION INITIATION
CC STEP (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39688; G1045764; -.
DR EMBL; U02212; G406492; -.
DR TIGR; MG087; -.
DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
DR PFAM; PF00164; S12; 1.
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 139 AA; 15641 MW; 695CE639 CRC32;
Query Match 40.9%; Score 47; DB 1; Length 139;
Best Local Similarity 45.0%; Pred. No. 1.7;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 2 IVTDGTVNDKNKPNQWVKY 21
: | ||: ||| :|||
Db 46 VCTRVGTMTPKPNSALRKY 65
RESULT 6
ID RS12_MYCPN STANDARD; PRT; 139 AA.
AC P75546;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE 30S RIBOSOMAL PROTEIN S12.
GN RPSL.
OS MYCOPLASMA PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE; 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERRMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL pneumoniae."
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC -1- FUNCTION: PROTEIN S12 IS INVOLVED IN THE TRANSLATION INITIATION
CC STEP (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000058; G1674308; -.

```

```
DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
DR PFAM; PF00164; S12; 1.
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 139 AA; 15629 MW; 507545FC CRC32;

Query Match 40.9%; Score 47; DB 1; Length 139;
Best Local Similarity 45.0%; Pred. No. 1.7;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 IVTADGTVNDKPNQWVRKY 21
DQ 46 VCTRVGTMTPKPKNSALRKY 65

RESULT 7
RS12_STAAU
ID RS12_STAAU STANDARD; PRT; 137 AA.
AC P48942;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE 305 RIBOSOMAL PROTEIN S12.
GN RPSL.
OS STAPHYLOCOCCUS AUREUS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
CC STAPHYLOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCTC 8325;
RA WADA A., WATANABE H.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: PROTEIN S12 IS INVOLVED IN THE TRANSLATION INITIATION
CC STEP (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
DR EMBL; U20869; G706921;
DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
DR PFAM; PF00164; S12; 1.
KW RIBOSOMAL PROTEIN.
SQ SEQUENCE 137 AA; 15287 MW; 716FF90D CRC32;

Query Match 40.9%; Score 47; DB 1; Length 137;
Best Local Similarity 45.0%; Pred. No. 1.7;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 IVTADGTVNDKPNQWVRKY 21
DQ 46 VCTRVGTMTPKPKNSALRKY 65

RESULT 8
CBF5_CANAL
ID CBF5_CANAL STANDARD; PRT; 479 AA.
AC O43101;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CENTROMERE/MICROTUBULE BINDING PROTEIN CBF5 (CENTROMERE-BINDING FACTOR
DE 5) (NUCLEOLAR PROTEIN CBF5).
GN CBF5.
OS CANDIDA ALBICANS (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
CC CANDIDACEAE; CANDIDA.
```

```

[1]
RN SEQUENCE FROM N.A.
RP JIANG W., CLIFFORD J., KOLTIN Y.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC SOME WAY ASSOCIATED WITH THE CBF3 110 KD SUBUNIT (CBF3A) (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
DR EMBL; U59149; G2737890;
KW MICROTUBULES; CENTROMERE; REPEAT; NUCLEAR PROTEIN; DNA-BINDING.
SQ SEQUENCE 479 AA; 54321 MW; 2E35229A CRC32;

Query Match 40.0%; Score 46; DB 1; Length 479;
Best Local Similarity 46.7%; Pred. No. 9.2;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 7 GTVNDKPNQWVRKY 21
DQ 373 GRVNTPEWKKDY 387

RESULT 9
VGL2_CVMA5
ID VGL2_CVMA5 STANDARD; PRT; 1324 AA.
AC P11224;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PELOMER PROTEIN)
DE [CONTAINS: SPIKE PROTEIN S1 (90B); SPIKE PROTEIN S2 (90A)].
GN S.
OS MURINE CORONAVIRUS MHV (STRAIN A59).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NIDOVIRALES;
OC CORONAVIRIDAE; CORONAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88072088.
RA LUTJES W., STURMAN L.S., BREDENBEEK P.J., CHARITE J.,
RA VAN DER ZELST B.A.M., HORZINEK M.C., SPAAN W.J.M.;
RT Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and
RT identification of the trypsin cleavage site."
RL VIROLOGY 161:479-487(1987).
CC -!- FUNCTION: THE PELOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC AND IN SYNCYTIIUM FORMATION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
DR EMBL; M18379; G331845;
DR PIR; A27402; VGH59.
KW GLYCOPROTEIN; ENVELOPE PROTEIN; TRANSMEMBRANE; SIGNAL.
FT SIGNAL 1 16
FT CHAIN 17 1324 SPIKE E2 GLYCOPROTEIN.
```

FT CHAIN 17 717 SPIKE PROTEIN S1 (90B).
 FT CHAIN 718 1324 SPIKE PROTEIN S2 (90A).
 FT DOMAIN 17 1265 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1266 1286 POTENTIAL.
 FT DOMAIN 1287 1324 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1287 1304 CYS-RICH.
 FT CARBOHYD 31 31 POTENTIAL.
 FT CARBOHYD 60 60 POTENTIAL.
 FT CARBOHYD 192 192 POTENTIAL.
 FT CARBOHYD 357 357 POTENTIAL.
 FT CARBOHYD 435 435 POTENTIAL.
 FT CARBOHYD 530 530 POTENTIAL.
 FT CARBOHYD 625 625 POTENTIAL.
 FT CARBOHYD 657 657 POTENTIAL.
 FT CARBOHYD 665 665 POTENTIAL.
 FT CARBOHYD 688 688 POTENTIAL.
 FT CARBOHYD 737 737 POTENTIAL.
 FT CARBOHYD 754 754 POTENTIAL.
 FT CARBOHYD 893 893 POTENTIAL.
 FT CARBOHYD 1180 1180 POTENTIAL.
 FT CARBOHYD 1190 1190 POTENTIAL.
 FT CARBOHYD 1209 1209 POTENTIAL.
 FT CARBOHYD 1225 1225 POTENTIAL.
 FT CARBOHYD 1246 1246 POTENTIAL.
 SQ SEQUENCE 1324 AA; 146019 MW; 82386829 CRC32;

Query Match 40.0%; Score 46; DB 1; Length 1324;
 Best Local Similarity 42.9%; Pred. No. 28;
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 8 TVNDKPNQWVKY 21
 I::: I: I: I:
 Db 437 TINHPSSWNR 450

RESULT 10
 ID LUXS_VIBFI STANDARD; PRT; 250 AA.
 AC P35327;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE LUXR REGULATORY PROTEIN.
 GN LUXR.
 OS VIBRIO FISCHERI.
 CC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIONACEAE; VIBRIO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ES114;
 RA GRAY K.M., GREENBERG E.P.;
 RT "Sequencing and analysis of luxR and luxI, the luminescence regulatory genes from the squid light organ symbiont *Vibrio fischeri* ES114.";
 RL MOL. MAR. BIOL. BIOTECHNOL. 1:414-419(1992).
 CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF THE BIOLUMINESCENCE OPERON. BINDS TO THE OHFL AUTOINDUCER.
 CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL REGULATORS. AUTOINDUCER REGULATED SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M96844; G295436; -
 DR PROSITE; PS00622; HTH_LUXR_FAMILY; 1.
 DR PFAM; PF00196; Gere; 1.
 KW TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR;
 KW AUTOINDUCER INDUCTION; LUMINESCENCE.
 FT DOMAIN 11 20 ARG/LYS-RICH (BASIC).

FT DOMAIN 64 67 ARG/LYS-RICH (BASIC).
 FT DOMAIN 104 114 LYS-RICH (BASIC).
 FT DOMAIN 185 188 ARG/LYS-RICH (BASIC).
 FT DNA_BIND 200 219 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 250 AA; 28519 MW; 0F9F6AAE CRC32;

Query Match 39.6%; Score 45.5; DB 1; Length 250;
 Best Local Similarity 42.9%; Pred. No. 5.4;
 Matches 9; Conservative 5; Mismatches 6; Indels 1; Gaps 1;

QY 1 AIVTADCTVNDKPNQWVKY 21
 I::: I: I: I:
 Db 50 SIIPDVSIIDNYPEKW-RKY 69

RESULT 11
 ID NATL_YEAST STANDARD; PRT; 853 AA.
 AC P12945;
 DT 01-JAN-1990 (REL. 13, CREATED)
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE N-TERMINAL ACETYLTRANSFERASE 1 (EC 2.3.1.88) (AMINO-TERMINAL, ALPHA-AMINO, ACETYLTRANSFERASE 1).
 GN NAT1 OR AAA1 OR YDL040C OR D2720.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 CC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
 CC SACCHAROMYCETACEAE; SACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE; 89308659.
 RA LEE F.-J.S., LIN L.-W., SMITH J.A.;
 RT "Molecular cloning and sequencing of a cDNA encoding N alpha-acetyltransferase from *Saccharomyces cerevisiae*.";
 RL J. BIOL. CHEM. 264:12339-12343(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 90005412.
 RA MULLEN J.R., KAYNE P.S., MOERSCHHELL R.P., TSUNASAWA S., GRIBSKOV M., COLAVITO-SHEPANSKI M., GRUNSTEIN M., SHERMAN F., STERNGLANZ R.;
 RT "Identification and characterization of genes and mutants for an N-terminal acetyltransferase from yeast.";
 RL EMBO J. 8:2067-2075(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA PAULIN L., SAREN A.M., LAAMANEN P.;
 RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- FUNCTION: N-TERMINAL ACETYLATION OF PROTEINS. N-ACETYLATION PLAYS A ROLE IN NORMAL EUKARYOTIC TRANSLATION AND PROCESSING. PROTECT AGAINST PROTEOLYTIC DEGRADATION AND PROTEIN TURNOVER. NAT1 ACTS ON PROTEINS WITH THE N-TERMINAL SERINE, GLYCINE OR ALANINE.
 CC -!- CATALYTIC ACTIVITY: ACETYL-COA + PEPTIDE -> COA + N-ACETYLPEPTIDE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: LIMITED, WITH OTHER ACETYLTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M23166; G172028; -
 DR EMBL; X15135; G4028; -
 DR EMBL; Z71781; E237293; -
 DR EMBL; Z74088; E252992; -
 DR PIR; S05783; XYBYTL.
 DR SGD; L0000002; NAT1.
 KW TRANSFERASE; ACYLTRANSFERASE; ACETYLATION.


```
FT INT_MET 0 0
SQ MOD_RES 1 1 ACETYLATION (POTENTIAL).
SQ SEQUENCE 853 AA; 98773 MW; EE9C0F59 CRC32;

Query Match 39.6%; Score 45.5; DB 1; Length 853;
Best Local Similarity 76.9%; Pred. No. 21;
Matches 10; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 10 NDNKNOW-VKRY 21
Db 537 NDIKENQMLVKRY 549

RESULT 12
BGLS_CALSA STANDARD; PRT; 455 AA.
AC P10482;
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE BETA-GLUCOSIDASE A (EC 3.2.1.21) (GENTIOBIASE) (CELLOBIASE) (BETA-D-
DE GLUCOSIDE GLUCOHYDROLASE) (AMYGALASE).
GN BGLA.
OS CALDCELLUM SACCHAROLYTICUM (CALDCELLULOSIRUPTOR SACCHAROLYTICUS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP;
OC THERMOANAEROBACTER GROUP; CALDCELLULOSIRUPTOR.
RN [1]
RP SEQUENCE FROM N.A.
RX LOVE D.R., BERGQUIST P.L.;
RX MEDLINE; 89127149.
RT "Sequence structure and expression of a cloned beta-glucosidase gene
RT from an extreme thermophile.";
RL MOL. GEN. GENET. 213:84-92(1988).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GLUCOSE RESIDUES WITH RELEASE OF BETA-D-GLUCOSE.
CC -1- CALDCELLUM SACCHAROLYTICUM IS A EXTREME THERMOPHILE AND APPEARS
CC TO BE A GRAM-POSITIVE ANAEROBIC BACTERIUM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X12575; G40644; -
DR PIR; S03813; S03813.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1.1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1.2; 1.
DR PFAM; PF00232; glycosyl_hydrol; 1.
DR HSSP; P11546; 1PBG.
KW HYDROLASE; GLYCOSIDASE; CELLULOSE DEGRADATION.
FT ACT_SITE 165 165 PROTON DONOR (POTENTIAL).
FT ACT_SITE 363 363 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 455 AA; 53492 MW; 6A58DD93 CRC32;

Query Match 39.1%; Score 45; DB 1; Length 455;
Best Local Similarity 40.0%; Pred. No. 12;
Matches 8; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

Qy 2 IVTADGTVDNKNPNQWVKRY 21
Db 371 IVTEDKVKHDSKRIEYLQKH 390

RESULT 13
CHLI_OLIU STANDARD; PRT; 334 AA.
ID CHLI_OLIU
AC Q32742;

FT INT_MET 0 0
SQ MOD_RES 1 1 ACETYLATION (POTENTIAL).
SQ SEQUENCE 853 AA; 98773 MW; EE9C0F59 CRC32;

Query Match 39.6%; Score 45.5; DB 1; Length 853;
Best Local Similarity 76.9%; Pred. No. 21;
Matches 10; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 10 NDNKNOW-VKRY 21
Db 537 NDIKENQMLVKRY 549

RESULT 12
BGLS_CALSA STANDARD; PRT; 455 AA.
AC P10482;
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE BETA-GLUCOSIDASE A (EC 3.2.1.21) (GENTIOBIASE) (CELLOBIASE) (BETA-D-
DE GLUCOSIDE GLUCOHYDROLASE) (AMYGALASE).
GN BGLA.
OS CALDCELLUM SACCHAROLYTICUM (CALDCELLULOSIRUPTOR SACCHAROLYTICUS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP;
OC THERMOANAEROBACTER GROUP; CALDCELLULOSIRUPTOR.
RN [1]
RP SEQUENCE FROM N.A.
RX LOVE D.R., BERGQUIST P.L.;
RX MEDLINE; 89127149.
RT "Sequence structure and expression of a cloned beta-glucosidase gene
RT from an extreme thermophile.";
RL MOL. GEN. GENET. 213:84-92(1988).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GLUCOSE RESIDUES WITH RELEASE OF BETA-D-GLUCOSE.
CC -1- CALDCELLUM SACCHAROLYTICUM IS A EXTREME THERMOPHILE AND APPEARS
CC TO BE A GRAM-POSITIVE ANAEROBIC BACTERIUM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X12575; G40644; -
DR PIR; S03813; S03813.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1.1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1.2; 1.
DR PFAM; PF00232; glycosyl_hydrol; 1.
DR HSSP; P11546; 1PBG.
KW HYDROLASE; GLYCOSIDASE; CELLULOSE DEGRADATION.
FT ACT_SITE 165 165 PROTON DONOR (POTENTIAL).
FT ACT_SITE 363 363 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 455 AA; 53492 MW; 6A58DD93 CRC32;

Query Match 39.1%; Score 45; DB 1; Length 455;
Best Local Similarity 40.0%; Pred. No. 12;
Matches 8; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

Qy 2 IVTADGTVDNKNPNQWVKRY 21
Db 371 IVTEDKVKHDSKRIEYLQKH 390

RESULT 13
CHLI_OLIU STANDARD; PRT; 334 AA.
ID CHLI_OLIU
AC Q32742;

DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MAGNESIUM-CHELATASE SUBUNIT CHLI (MG-PROTOPORPHYRIN IX CHELATASE).
GN CHLI OR CCSA.
OS OLISTHODISCUS LUTEUS (HETEROSIGMA AKASHIMO).
OC CHLOROPLAST.
OC EUKARYOTA; STRAMENOPILES; RAPHIIDOPHYCEAE; OLISTHODISCUS.
RN [1]
RP SEQUENCE FROM N.A.
RA VALENTIN K.U., DARBY C., CATTOLICO R.A.;
RL SUBMITTED (MAR-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS; INTRODUCES
CC A MAGNESIUM ION INTO PROTOPORPHYRIN IX TO YIELD MG-
CC PROTOPORPHYRIN IX.
CC -1- PATHWAY: CHLOROPHYLL BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE CHLI/BCHI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z21959; G288236; -
DR PFAM; PF01078; Mg_chelatase; 1.
DR MENDEL; 2292; OLILU:CHLI.1.
KW PHOTOSYNTHESIS; CHLOROPHYLL BIOSYNTHESIS; CHLOROPLAST;
KW ATP-BINDING.
FT NP_BIND 43 50 ATP (POTENTIAL).
SQ SEQUENCE 334 AA; 37208 MW; F47E0842 CRC32;

Query Match 38.3%; Score 44; DB 1; Length 334;
Best Local Similarity 36.8%; Pred. No. 13;
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 3 VTADGTVDNKNPNQWVKRY 21
Db 211 VVEERTFLDLNPEEWINKY 229

RESULT 14
RS12_STRPN STANDARD; PRT; 137 AA.
ID RS12_STRPN
AC P30891;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE 30S RIBOSOMAL PROTEIN S12.
GN RPSL OR STR.
OS STREPTOCOCCUS PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R800.
RX MEDLINE; 93096605.
RA SALLES C., CREANCIER L., CLAVERYS J.-P., MEJEAN V.;
RT "The high level streptomycin resistance gene from Streptococcus
RT pneumoniae is a homologue of the ribosomal protein S12 gene from
RT Escherichia coli.";
RL NUCLEIC ACIDS RES. 20:6103-6103(1992).
CC -1- FUNCTION: PROTEIN S12 IS INVOLVED IN THE TRANSLATION INITIATION
CC STEP (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S12P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

Search completed: November 13, 1999, 10:33:39
Job time: 5189 sec

DR EMBL: Z15120; G47453; -.
DR PIR: S26680; S26680.
DR PROSITE: PS00055; RIBOSOMAL_S12; 1.
DR PFAM: PF00164; S12; 1.
KW RIBOSOMAL PROTEIN; ANTIBIOTIC RESISTANCE.
FT VARIANT 56 K -> T (IN STR41 MUTATION, HIGH LEVEL
OF RESISTANCE TO STREPTOMYCIN).
SQ SEQUENCE 137 AA; 15144 MW; 98659752 CRC32;

Query Match 38.3%; Score 44; DB 1; Length 137;
Best Local Similarity 40.0%; Pred. No. 4.8;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 2 IVTADGTVDNKNPQWVRKY 21
: | ||| :||| :|||
Db 46 VATRYGTMTPKPKNSALRKF 65

RESULT 15

MTHL_HAEIN
ID MTHL_HAEIN STANDARD; PRT; 359 AA.
AC P20590;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE MODIFICATION METHYLASE HINFI (EC 2.1.1.72) (ADENINE-SPECIFIC
DE METHYLTRANSFERASE HINFI) (M.HINFI).
GN HINFI.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RF;
RX MEDLINE; 89108022.
RA CHANDRASEGARAN S., LUNNEN K.D., SMITH H.O., WILSON G.G.;
RT "Cloning and sequencing the HinfI restriction and modification
RT genes";
RL GENE 70:387-392(1988).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC GANTC, CAUSES SPECIFIC METHYLATION ON A-2 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE HINFI ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA ADENINE -
CC S-ADENOSYL-L-HOMOCYSTEINE + DNA 6-METHYLAMINOPURINE.
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----

DR EMBL: M22862; G148945; -.
DR PIR: JT0391; JT0391.
DR REBASE: RB00028; HinfI.
DR PROSITE: PS00092; NC_MTASE; 1.
KW TRANSFERASE; METHYLTRANSFERASE; RESTRICTION SYSTEM.
SQ SEQUENCE 359 AA; 41805 MW; BAEDCB0B CRC32;

Query Match 37.4%; Score 43; DB 1; Length 359;
Best Local Similarity 47.4%; Pred. No. 20;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 2 IVTADGTVDNKNPQWVRK 20
||| ||| :||| :|||
Db 303 IVTQGNVKNDEETLSIHK 321

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:31 ; Search time 139.86 Seconds
(without alignments)
9.241 Million cell updates/sec

Title: US-08-913-430-9

Perfect score: 115

Sequence: 1 AIVTADGTVDNKNPNQVRKY 21

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPTREMBL10.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_protist.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	42.6	240	P74946	P74946 vibrio angu
2	49	42.6	312	O56666	O56666 duck enteri
3	48	41.7	1360	O55253	O55253 murine hepa
4	47	40.9	1361	O83331	O83331 murine hepa
5	47	40.9	1361	P90210	P90210 murine hepa
6	46.5	40.4	266	O48979	O48979 hordeum vul
7	46	40.0	467	O65550	O65550 arabidopsis
8	46	40.0	1324	O39227	O39227 murine hepa
9	46	40.0	539	O9YQSO	O9YQSO human calic
10	45	39.1	316	Q46064	Q46064 corynebacte
11	45	39.1	539	Q68537	Q68537 human calic
12	44	38.3	117	O51021	O51021 neisseria g
13	43	37.4	747	O69735	O69735 mycobacteri
14	43	37.4	6260	O54299	O54299 streptomyce
15	43	37.4	744	O42958	O42958 schizosacch
16	43	37.4	1089	O75137	O75137 homo sapien
17	43	37.4	451	O94487	O94487 dictyosteli
18	43	37.4	1376	O83330	O83330 murine hepa
19	42.5	37.0	436	O15343	O15343 homo sapien
20	42.5	37.0	537	O77987	O77987 human immun
21	42	36.5	443	O67166	O67166 aquifex aeo
22	42	36.5	662	O25886	O25886 helicobacte
23	42	36.5	1402	O86335	O86335 mycobacteri
24	42	36.5	521	O87767	O87767 lactococcus
25	42	36.5	1518	O21850	O21850 caenorhabdi
26	42	36.5	255	O18435	O18435 helicoverpa
27	42	36.5	1293	O04251	O04251 arabidopsis
28	42	36.5	539	O66287	O66287 unidentified
29	42	36.5	344	O86899	O86899 equine infl

30	42	36.5	579	12	Q84396	Q84396 porcine par
31	41.5	36.1	233	12	O55578	O55578 leucania se
32	41.5	36.1	533	12	Q78676	Q78676 human immun
33	41	35.7	211	2	P71189	P71189 escherichia
34	41	35.7	328	3	O13957	O13957 schizosacch
35	41	35.7	804	4	O95294	O95294 homo sapien
36	41	35.7	70	5	O94967	O94967 drosophila
37	41	35.7	429	5	Q16994	Q16994 anthopleura
38	41	35.7	166	10	O49360	O49360 arabidopsis
39	41	35.7	427	10	P93185	P93185 hordeum vul
40	41	35.7	208	10	Q41230	Q41230 nicotiana t
41	41	35.7	211	10	Q41231	Q41231 nicotiana t
42	41	35.7	427	10	P93187	P93187 hordeum vul
43	41	35.7	503	10	Q92Q37	Q92Q37 arabidopsis
44	41	35.7	835	12	O09725	O09725 human calic
45	41	35.7	863	12	O93124	O93124 human calic

ALIGNMENTS

RESULT 1

P74946 PRELIMINARY: PRT: 240 AA.

AC P74946; DT 01-FEB-1997 (TREMREL. 02, Created)

DT 01-FEB-1997 (TREMREL. 02, Last sequence update)

DT 01-NOV-1998 (TREMREL. 08, Last annotation update)

DE VANR.

GN VANR.

OS Vibrio anguillarum.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

[1]

RN SEQUENCE FROM N.A.

RC STRAIN-NB10;

RX MEDLINE; 97284506.

RA MILTON D.L., HARDMAN A., CAMARA M., CHABRA S.R., BYCROFT B.W., STEWART G.S.A.B., WILLIAMS P.;

RT "Quorum sensing in Vibrio anguillarum: characterization of the vanI/vanR locus and identification of the autoinducer N-(3-oxodecanoyl)-L-homoserine lactone.";

RL J. Bacteriol. 179:3004-3012(1997).

DR EMBL; O69677; AAC45213.1; -.

DR PFAM; PF00196; GcrE; 1.

SQ SEQUENCE 240 AA; 27253 MW; 559D911E CRC32;

Query Match 42.6%; Score 49; DB 2; Length 240;

Best Local Similarity 44.4%; Pred. No. 3.3;

Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 TADGTVDNKNPNQVRKY 21

Db 48 TSETLVDNPNQVRKY 65

RESULT 2

O56666 PRELIMINARY: PRT: 312 AA.

AC O56666; DT 01-JUN-1998 (TREMREL. 06, Created)

DT 01-JUN-1998 (TREMREL. 06, Last sequence update)

DT 01-JUN-1998 (TREMREL. 06, Last annotation update)

DE UL6 AND UL7 GENES PARTIAL CDS (FRAGMENT).

OS duck enteritis virus.

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

[1]

RN SEQUENCE FROM N.A.

RP PLUMMER P.J., ALEFANTIS T., KAPLAN S.L., O'CONNELL P., SHAWKY S., SCHAT K.A.;

RA Avian Dis. 0:0-0(1998).

DR EMBL; AF043730; AAC03544.1; -.

FT NON_TER 312 312
SQ SEQUENCE 312 AA; 34899 MW; E31741B1 CRC32;

Query Match 42.6%; Score 49; DB 12; Length 312;
Best Local Similarity 47.4%; Pred. No. 4.4;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 AIVTADGTVNDKPNQWVR 19
|||||: : : : :
DB 280 AIVTADGVLNEALKCEWIR 298

RESULT 3
O55253 PRELIMINARY; PRT; 1360 AA.
AC O55253;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE SPIKE GLYCOPROTEIN.
OS murine hepatitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DVIM;
RA SUGIYAMA K., MORITA E., EBINA H., YOKOYAMA M., MUTO A., HIMENO H.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008940; BAA23719.1; -
SQ SEQUENCE 1360 AA; 149337 MW; CDB47FDC CRC32;

Query Match 41.7%; Score 48; DB 12; Length 1360;
Best Local Similarity 33.3%; Pred. No. 31;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 AIVTADGTVNDKPNQWVRKY 21
::: ||||: | : | : |
DB 429 SLAQSNVTNNHPSSWNRKY 449

RESULT 4
Q83331 PRELIMINARY; PRT; 1361 AA.
AC Q83331;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE S GLYCOPROTEIN.
OS murine hepatitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE; 95304829.
RA KUNITA S., ZHANG L., HOMBERGER F.R., COMPTON S.R.;
RT "Molecular characterization of the S proteins of two enterotropic
murine coronavirus strains";
RL Virus Res. 35:277-289(1995).
DR EMBL; U14646; AAA87063.1; -
SQ SEQUENCE 1361 AA; 150069 MW; 71B8E137 CRC32;

Query Match 40.9%; Score 47; DB 12; Length 1361;
Best Local Similarity 50.0%; Pred. No. 45;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 8 TVNDKPNQWVRKY 21
||||: | : | : |
DB 437 TVNNHPSSWNRKY 450

RESULT 5
P90210 PRELIMINARY; PRT; 1361 AA.
AC P90210;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE SPIKE PROTEIN.
OS murine hepatitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MHV-2;
RX MEDLINE; 97167535.
RA YAMADA Y.K., TAKIMOTO K., YABE M., TAGUCHI F.;
RT "Acquired fusion activity of a murine coronavirus MHV-2 variant with
mutations in the proteolytic cleavage site and the signal sequence of
the S protein";
RL Virology 227:215-219(1997).
DR EMBL; U72635; AAC56567.1; -
SQ SEQUENCE 1361 AA; 149841 MW; E6CCF82D CRC32;

Query Match 40.9%; Score 47; DB 12; Length 1361;
Best Local Similarity 50.0%; Pred. No. 45;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 8 TVNDKPNQWVRKY 21
||||: | : | : |
DB 437 TVNNHPSSWNRKY 450

RESULT 6
O48979 PRELIMINARY; PRT; 266 AA.
AC O48979;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE NBS-LRR TYPE RESISTANCE PROTEIN (FRAGMENT).
GN B8.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Hordeum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. INGRID;
RX MEDLINE; 98081880.
RA LEISTER D., KURTH J., LAURIE D.A., YANO M., SASAKI T., DEVOS K.,
GRANER A., SCHULZE-LEFERT P.;
RT "Rapid reorganization of resistance gene homologues in cereal
genomes";
RL Proc. Natl. Acad. Sci. U.S.A. 95:370-375(1998).
DR EMBL; AF032686; AAB96983.1; -
DR PFAM; PF00931; NB-ARC; 1.
FT NON_TER 1
FT NON_TER 266
SQ SEQUENCE 266 AA; 30026 MW; 62CF4FE1 CRC32;

Query Match 40.4%; Score 46.5; DB 10; Length 266;
Best Local Similarity 55.6%; Pred. No. 9.2;
Matches 10; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 AIVTADG-TVNDKPNQW 17
||||: | : | : |
DB 169 AVTVAGLLVNOGKPEQW 186

RESULT 7

065550
ID O65550 PRELIMINARY; PRT; 467 AA.
AC O65550;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
DE PUTATIVE CALMODULIN-BINDING PROTEIN.
GN F6118.90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
[1]
RN SEQUENCE FROM N.A.
RA BEVAN M., BENES V., RECHMANN S., BORKOVA D., ANSORGE W., HOEISEL J.,
RA MEVES H.W., MAYER K., SCHUELLER C.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA EU ARABIDOPSIS SEQUENCING PROJECT;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL022198; CAAL18193.1; -;
SQ SEQUENCE 467 AA; 52831 MW; 132A5733 CRC32;

Query Match 40.0%; Score 46; DB 10; Length 467;
Best Local Similarity 40.0%; Pred. No. 20;
Matches 8; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 2 IVTADGTVDNKNPNQWRKY 21
| | | | | | | | | |
Db 49 IVVLGDGFDEDDKDWTRH 68

RESULT 8
039227 PRELIMINARY; PRT; 1324 AA.
ID O39227;
AC O39227;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JAN-1998 (Tremblrel. 05, Last annotation update)
DE E2 GLYCOPROTEIN PRECURSOR.
OS murine hepatitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-MHV-A59;
RC LEPARC-GOFFART I., HINGLEY S.T., CHUA M.M., JIANG X., LAVI E.,
RA WEISS S.R.;
RL Virology 0:0-0(1997).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-MHV-A59;
RC WEISS S.R., LEPARC-GOFFART I., HINGLEY S.T.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF029248; AAB86819.1; -;
SQ SEQUENCE 1324 AA; 145899 MW; EF34B05C CRC32;

Query Match 40.0%; Score 46; DB 12; Length 1324;
Best Local Similarity 42.9%; Pred. No. 62;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 8 TVNDKNPNQWRKY 21
| | | | | | | | | |
Db 437 TINNPSPSWNRRY 450

RESULT 9
Q9YQSO PRELIMINARY; PRT; 539 AA.
ID Q9YQSO

Q9YQSO;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE CAPSID PROTEIN.
OS human calicivirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Calicivirus.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-GRIMSBY;
RA GREEN J., VINJE J., GALLIMORE C.I., KOOPMANS M., BROWN D.W.;
RL "Capsid protein diversity among Genogroup I and II human
RL caliciviruses";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ004864; CAA06169.1; -;
SQ SEQUENCE 539 AA; 58879 MW; 0AA0C6F0 CRC32;

Query Match 40.0%; Score 46; DB 12; Length 539;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 3 VTADGTVDNKNPNQWRKY 18
| | | | | | | | | |
Db 388 VIODGNHNEPQQWV 403

RESULT 10
Q46064 PRELIMINARY; PRT; 316 AA.
ID Q46064;
AC Q46064;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE ORF3 AND AROP GENE.
OS Corynebacterium glutamicum.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
OC Corynebacterium.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-ATCC 13032;
RX MEDLINE; 96011392.
RA WEHRMANN A., MORAKKABATI S., KRAEMER R., SAHM H., EGGLING L.;
RL "Functional analysis of sequences adjacent to dape of Corynebacterium
RL glutamicum reveals the presence of aroP, which encodes the aromatic
RL amino acid transporter";
RL J. Bacteriol. 177:5991-5993(1995).
DR EMBL: X85965; CAA59949.1; -;
SQ SEQUENCE 316 AA; 34948 MW; 73DCF194 CRC32;

Query Match 39.1%; Score 45; DB 2; Length 316;
Best Local Similarity 31.0%; Pred. No. 19;
Matches 9; Conservative 4; Mismatches 8; Indels 8; Gaps 1;

Qy 1 AIVTADGTVDN-----KPNQWRKY 21
| | | | | | | | | |
Db 14 ANIAMDGTILDTWYPEQIFNPQWERY 42

RESULT 11
Q68537 PRELIMINARY; PRT; 539 AA.
ID Q68537;
AC Q68537;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE CAPSID.
OS human calicivirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Calicivirus.

RN SEQUENCE FROM N.A.
 RP STRAIN-CAMBERWELL;
 RX MEDLINE; 96310230.
 RA CAUCHI M.R., DOULTREE J.C., MARSHALL J.A., WRIGHT P.J.;
 RT "Molecular characterization of Camberwell virus and sequence
 RL variation in ORF3 of small round-structured (Norwalk-like) viruses.";
 J. Med. Virol. 49:70-76(1996).
 DR EMBL; U46500; AAC54893.1; -.
 DR PFAM; PF00915; Calici_coat.1.
 SQ SEQUENCE 539 AA; 58869 MW; 23BB149B CRC32;

Query Match 39.1%; Score 45; DB 12; Length 539;

Best Local Similarity 50.0%; Pred. No. 34;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 VTADGTVDNKNQVW 18
 | | | | | | | | | |
 DB 388 VIQGDGHQNEPQVW 403

RESULT 12

ID Q51021 PRELIMINARY; PRT; 117 AA.
 AC Q51021;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE FILE1 L-PILIN (MS11-D1).
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MS11;
 RX MEDLINE; 91312130.
 RA MANNING P.A., KAUFMANN A., ROLL U., POHLNER J., MEYER T.F., HAAS R.;
 RT "L-pilin variants of Neisseria gonorrhoeae MS11";
 RL Mol. Microbiol. 5:917-926(1991).
 DR EMBL; X58403; CAA41302.1; -.
 SQ SEQUENCE 117 AA; 12495 MW; AA6550A4 CRC32;

Query Match 38.3%; Score 44; DB 2; Length 117;

Best Local Similarity 81.8%; Pred. No. 9.3;

Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VTADGTVDNKN 13
 | | | | | | | | | |
 DB 89 VTADGTGNDGK 99

RESULT 13

ID O69735 PRELIMINARY; PRT; 747 AA.
 AC O69735;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE HYPOTHETICAL 80.9 KD PROTEIN.
 GN MTV027.05.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA BADCOCK K., CHURCHER C.M.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;

Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RL SEQUENCE FROM N.A.
 RN STRAIN-H37RV;
 RC MEDLINE; 96181548.
 RA PHILIPP W.J., POULTER S., EIGMEIER K., PASCOPELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RT "An integrated map of the genome of the tubercle bacillus,
 RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
 RT leprae";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
 DR EMBL; AL021210; CAA17962.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 747 AA; 80913 MW; 3030D433 CRC32;

Query Match 37.4%; Score 43; DB 2; Length 747;

Best Local Similarity 50.0%; Pred. No. 99;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 VTADGTVDNKNQVWK 20
 | | | | | | | | | |
 DB 356 VHEGDTADDRLPNQSFQ 373

RESULT 14

ID Q54299 PRELIMINARY; PRT; 6260 AA.
 AC Q54299;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE POLYKETIDE SYNTHASE.
 GN RAPC.
 OS Streptomyces hygroscopicus.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL 5491;
 RX MEDLINE; 95372374.
 RA SCHWECHE T., APARICIO J.F., MOLNAR I., KOENIG A., KHAW L.E.,
 RA HAYDOCK S.F., OLYNYK M., CAFFEY P., CORTES J., LESTER J.B.,
 RA BOEHM G.A., STAUNTON J., LEADLAY P.F.;
 RT "The biosynthetic gene cluster for the polyketide immunosuppressant
 RT rapamycin";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL 5491;
 RA MOLNAR I., APARICIO J.F., HAYDOCK S.F., EE KHAW L., SCHWECHE T.,
 RA KOENIG A., STAUNTON J., LEADLAY P.F., STAUNTON J., LEADLAY P.F.;
 RL Gene 0:0-0(0).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL 5491;
 RA APARICIO J.F., MOLNAR I., SCHWECHE T., KOENIG A., HAYDOCK S.F.,
 RA EE KHAW L., STAUNTON J., LEADLAY P.F., LESTER J.B., BOEHM G.A.,
 RA STAUNTON J., LEADLAY P.F.;
 RL Gene 0:0-0(0).
 [4]
 DR EMBL; X86780; CAA60462.1; -.
 DR PFAM; PF00698; Acyl_transf. 4.
 DR PFAM; PF00107; adh_zinc. 1.
 DR PFAM; PF00109; ketoacyl-synt. 4.
 DR PFAM; PF00550; pp-binding. 4.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 4.
 DR PROSITE; PS01162; QOR_ZETA_CRYSTAL; 1.
 KW Transferase.
 SQ SEQUENCE 6260 AA; 659524 MW; 99FDECC3 CRC32;

Query Match 37.4%; Score 43; DB 2; Length 6260;
 Best Local Similarity 61.5%; Pred. No. 9.9e+02;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 IVTADGTVNDNKP 14
 :|||:||||:
 Db 4561 VVVALGVNDNRP 4573

RESULT 15
 O42958
 ID 042958 PRELIMINARY; PRT; 744 AA.
 AC 042958;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
 DE HYPOTHETICAL 82.9 KD PROTEIN.
 GN SPC1967.10C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA LYNE M., RAJANDREAM M.A., BARRELL B.G., DEVLIN K., CHURCHER C.M.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL021839; CAAL7064.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 744 AA; 82920 MW; A1B4FA45 CRC32;

Query Match 37.4%; Score 43; DB 3; Length 744;
 Best Local Similarity 43.8%; Pred. No. 99;
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 2 IVTADGTVNDNKPQW 17
 :|||||:
 Db 641 ILKQSGTVNQNEKEW 656

Search completed: November 13, 1999, 12:55:34
 Job time: 3033 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:59:55 ; Search time 104.22 Seconds
(without alignments)
3.409 Million cell updates/sec

Title: US-08-913-430-10

Perfect score: 60

Sequence: 1 MKLAKLLKGFXXXIK 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	90.0	15	1 W01028	Mycoplasma 60-64 k
2	36	60.0	205	1 R65007	Mouse tissue inhib
3	34	56.7	542	1 W79072	Sinapis alba cytoc
4	32	53.3	207	1 R65003	Cattle tissue inhi
5	32	53.3	207	1 R65004	Pig tissue inhib
6	31	51.7	207	1 P60786	Sequence of tissue
7	31	51.7	207	1 P60275	Sequence of human
8	31	51.7	207	1 P60592	Sequence of a huma
9	31	51.7	207	1 P60593	Sequence of a gibb
10	31	51.7	18	1 R20981	Sequence of amphip
11	31	51.7	18	1 R22834	Amphiphilic peptid
12	31	51.7	724	1 R29002	Protein derived fr
13	31	51.7	207	1 R65005	Human tissue inhib
14	31	51.7	437	1 W20294	H. pylori cell env
15	31	51.7	448	1 W20656	H. pylori transpor
16	31	51.7	207	1 W30309	Human TIMP-1. Huma
17	31	51.7	108	1 W61220	Streptococcus pneu
18	30	50.0	756	1 R08405	Amf105L encoded by
19	30	50.0	17	1 R21385	Sequence of amphip
20	30	50.0	17	1 R21386	Sequence of amphip
21	30	50.0	17	1 R22869	Amphiphilic peptid
22	30	50.0	17	1 R22870	Amphiphilic peptid
23	30	50.0	298	1 R36678	Ethylene-induced a
24	30	50.0	207	1 R65008	Mouse tissue inhib
25	30	50.0	1726	1 W00385	Truncated Plasmodi
26	30	50.0	179	1 W98674	H. pylori GHPO 532
27	30	50.0	98	1 Y11662	Human 5' EST seque
28	29	48.3	38	1 R07743	Lytic peptide with
29	29	48.3	586	1 R07274	Reverse transcript
30	29	48.3	984	1 P80927	Sequence of the hu
31	29	48.3	586	1 R1920	Reverse transcript
32	29	48.3	21	1 R31090	C-terminal substd.
33	29	48.3	21	1 R35308	Amphiphilic peptid
34	29	48.3	21	1 R36319	Amphiphilic ion ch
35	29	48.3	21	1 R39022	Biologically activ
36	29	48.3	21	1 R45060	Basic (positively
37	29	48.3	21	1 R50490	Amphiphilic peptid
38	29	48.3	21	1 R50373	Amphiphilic peptid
39	29	48.3	21	1 R50374	Amphiphilic peptid
40	29	48.3	21	1 R55914	Ion channel formin
41	29	48.3	21	1 R55629	Peptide which neut
42	29	48.3	21	1 R89992	Cancer treating, a
43	29	48.3	206	1 R65006	Rabbit tissue inhi

44 29 48.3 586 1 R73935 msDNA-Ec67 reverse
45 29 48.3 355 1 W82385 Flea saliva protei

ALIGNMENTS

RESULT 1
W01028
ID W01028 standard; Peptide; 15 AA.
AC W01028;
DT 19-JAN-1997 (first entry)
DE Mycoplasma 60-64 kDa protective antigen N-terminal peptide.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumoniae;
KW diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
FH Key Location/Qualifiers
FT misc_difference 11 /note= "undetermined amino acid"
FT misc_difference 12 /label= Asn, Leu
FT misc_difference 13 /label= Met, Val
FT misc_difference 13 /label= Met, Val
PN W09628472-Al.
PD 19-SEP-1996.
PR 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (UYME) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
DR WPI; 96-433763/43.
PT Putative protective antigens against Mycoplasma - used for the
PT detection, prevention or treatment of Mycoplasma infections, esp. M.
PT hyopneumoniae in swine
PS Claim 9; Page 27; 43pp; English.
CC Two 60-64 kDa putative protective antigens against Mycoplasma
CC were identified that respectively contain the N-terminal sequences
CC given in W01028 and W01029. The antigens were isolated from
CC Mycoplasma hyopneumoniae cells using antibody probes enriched with
CC Mycoplasma-specific antibodies. Other protective antigens were
CC also identified (see also W01024-27 and W01030-37). Protective
CC antigens and antibodies can be used in vaccines for preventing or
CC treating mycoplasma infections, partic. M. hyopneumoniae
CC infections in swine. They can also be used for diagnosis.
SQ Sequence 15 AA;

Query Match 90.0%; Score 54; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKLAKLLKGFXXXIK 15
DB 1 MKLAKLLKGFXXXIK 15

RESULT 2
R65007
ID R65007 standard; Protein; 205 AA.
AC R65007;
DT 24-OCT-1995 (first entry)
DE Mouse tissue inhibitor of metalloproteinase (TIMP-1).
KW Tissue inhibitor of metalloproteinase; diagnostic; therapeutic;
KW prophylaxis.
OS Mus musculus.
PN W09505478-A.
PD 23-FEB-1995.
PR 12-AUG-1994; U09188.
PR 12-AUG-1993; US-105263.
PR 13-DEC-1993; US-167463.
PA (REGC) UNIV CALIFORNIA.
PI Hawkes SP, Kishnani NS, Yang T;
DR WPI; 95-098775/13.
PT New human tissue inhibitor of metalloproteinase-3 - used to

PT develop prods. for diagnosis, therapy or prophylaxis of
 PS conditions with unwanted matrix metallo:proteinase activity.
 CC Disclosure; Fig 6A-B; 87pp; English.
 CC The protein sequence of the mouse TIMP-1 is compared with those
 CC from pig, human, rabbit, cattle and a 2nd mouse TIMP-1, human,
 CC mouse and cattle TIMP-2 and chicken (Ch) and mouse TIMP-3. A
 CC probe based on the CHIMP-3 amino acid sequence (R65000) is used
 CC to isolate DNA encoding human TIMP-3 from a human cDNA library.
 CC Human TIMP-3 can be used for the diagnosis, therapy or
 CC prophylaxis of conditions characterized by excess or unwanted
 CC matrix metalloproteinase activity, e.g. neoplasias, tumor
 CC metastasis, inflammatory disorders such as rheumatoid arthritis,
 CC ulcerations, reaction to infection, periodontal disease or
 CC osteoporosis. It can also be used in drug screening/design.
 SQ Sequence 205 AA;

Query Match 60.0%; Score 36; DB 1; Length 205;
 Best Local Similarity 60.0%; Pred. No. 6.5;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
 DB 64 IKMTKMLKGF 73

RESULT 3
 W79072 ID W79072 standard; Protein; 542 AA.
 AC W79072; 14-DEC-1998 (first entry)
 DE Sinapis alba cytochrome P450 monooxygenase.
 KW Cytochrome P450 monooxygenase; P450ox; Sorghum bicolor (L.) Moench;
 KW Sinapis alba; biosynthetic conversion; aloxime; nitrile; cyanohydrin;
 KW cyanogenic glycoside; transgenic plant; resistance.
 OS Sinapis alba.
 FH Key Location/Qualifiers
 FT Misc_difference 475 /note= "unspecified"
 FT W09840470-A2.
 PD 17-SEP-1998.
 PF 05-MAR-1998; E01253.
 PR 08-DEC-1997; EP-810954.
 PR 07-MAR-1997; EP-810132.
 PA (NOVS) NOVARTIS AG.
 PA (UYRO-) UNIV ROYAL VETERINARY & AGRIC.
 PI Bak S, Halkier BA, Kahn RA, Moeller BL;
 DR WPI; 98-320808/44.
 PT Cytochrome P450 monooxygenase of the cyanogenic glycoside pathway -
 PT useful for the production of plants with improved nutritive value or
 PT pest resistance
 PS Claim 28; Page 51-53; 32pp; English.
 CC The present sequence represents a cytochrome P450 monooxygenase from
 CC Sinapis alba. Cytochrome P450 monooxygenase catalyses: (i) the
 CC conversion of aloxime to a nitrile; and (ii) the nitrile to the
 CC corresponding cyanohydrin. DNA encoding cytochrome P450 monooxygenase
 CC can be used to obtain transgenic plants, for the purpose of improving
 CC the nutritive value or pest resistance of the plant. Cytochrome P450
 CC monooxygenase catalyses the conversion of aloximes to nitriles to
 CC cyanohydrins, which are the precursors of toxic cyanogenic glycosides,
 CC so staple food such as cassava and lima beans, as well as animal feed
 CC such as white clover, can be rendered less toxic by blocking the
 CC cytochrome P450 monooxygenase activity. Introducing the enzyme to
 CC plants or to certain tissues could help reduce crop damage since the
 CC product is also toxic to insects, acarids and nematodes.
 SQ Sequence 542 AA;

Query Match 56.7%; Score 34; DB 1; Length 542;
 Best Local Similarity 70.0%; Pred. No. 44;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10

DB 490 MLLARLLQGF 499

RESULT 4
 R65003 ID R65003 standard; Protein; 207 AA.
 AC R65003; 23-OCT-1995 (first entry)
 DE Cattle tissue inhibitor of metalloproteinase (TIMP-1).
 KW Tissue inhibitor of metalloproteinase; diagnostic; therapeutic;
 KW prophylaxis.
 OS Bos primigenius taurus.
 PN W09505478-A.
 PD 23-FEB-1995.
 PF 12-AUG-1994; U09188.
 PR 12-AUG-1993; US-105263.
 PR 13-DEC-1993; US-167463.
 PA (REGC) UNIV CALIFORNIA.
 PI Hawkes SP, Kishnani NS, Yang T;
 DR WPI; 95-098775/13.
 PT New human tissue inhibitor of metallo:proteinase-3 - used to
 PT develop prods. for diagnosis, therapy or prophylaxis of
 PT conditions with unwanted matrix metallo:proteinase activity.
 PS Disclosure; Fig 6A-B; 87pp; English.
 CC The protein sequence of the cattle TIMP-1 is compared with those
 CC from pig, human, rabbit and mouse TIMP-1, human, mouse and cattle
 CC TIMP-2 and chicken (Ch) and mouse TIMP-3. A probe based on the
 CC CHIMP-3 amino acid sequence (R65000) is used to isolate DNA
 CC encoding human TIMP-3 from a human cDNA library. Human TIMP-3
 CC can be used for the diagnosis, therapy or prophylaxis of
 CC conditions characterized by excess or unwanted matrix
 CC metalloproteinase activity, e.g. neoplasias, tumor metastasis,
 CC inflammatory disorders such as rheumatoid arthritis, ulcerations,
 CC reaction to infection, periodontal disease or osteoporosis. It can
 CC also be used in drug screening/design.
 SQ Sequence 207 AA;

Query Match 53.3%; Score 32; DB 1; Length 207;
 Best Local Similarity 50.0%; Pred. No. 40;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
 DB 63 IKMTKMLKGF 72

RESULT 5
 R65004 ID R65004 standard; Protein; 207 AA.
 AC R65004; 23-OCT-1995 (first entry)
 DE Pig tissue inhibitor of metalloproteinase (TIMP-1).
 KW Tissue inhibitor of metalloproteinase; diagnostic; therapeutic;
 KW prophylaxis.
 OS Sus scrofa.
 PN W09505478-A.
 PD 23-FEB-1995.
 PF 12-AUG-1994; U09188.
 PR 12-AUG-1993; US-105263.
 PR 13-DEC-1993; US-167463.
 PA (REGC) UNIV CALIFORNIA.
 PI Hawkes SP, Kishnani NS, Yang T;
 DR WPI; 95-098775/13.
 PT New human tissue inhibitor of metallo:proteinase-3 - used to
 PT develop prods. for diagnosis, therapy or prophylaxis of
 PT conditions with unwanted matrix metallo:proteinase activity.
 PS Disclosure; Fig 6A-B; 87pp; English.
 CC The protein sequence of the pig TIMP-1 is compared with those
 CC from cattle, human, rabbit and mouse TIMP-1, human, mouse and cattle
 CC TIMP-2 and chicken (Ch) and mouse TIMP-3. A probe based on the
 CC CHIMP-3 amino acid sequence (R65000) is used to isolate DNA

CC encoding human TIMP-3 from a human cDNA library. Human TIMP-3
 CC can be used for the diagnosis, therapy or prophylaxis of
 CC conditions characterized by excess or unwanted matrix
 CC metalloproteinase activity, e.g. neoplasias, tumor metastasis,
 CC inflammatory disorders such as rheumatoid arthritis, ulcerations,
 CC reaction to infection, periodontal disease or osteoporosis. It can
 CC also be used in drug screening/design.
 SQ Sequence 207 AA;

Query Match 53.3%; Score 32; DB 1; Length 207;
 Best Local Similarity 50.0%; Pred. No. 40;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
 :|: |: |||
 Db 63 IKTKMKYKGF 72

RESULT 6

ID P60786 standard; Protein; 207 AA.
 AC P60786;
 DT 08-AUG-1991 (first entry)
 DE Sequence of tissue inhibitor of metalloproteinase (TIMP).
 KW Connective tissue; extracellular matrix.
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT peptide 1..23
 FT /label= signal
 FT protein 24..207
 FT modified_site 53..55
 FT /label= potential N-glycosylation site
 FT modified_site 101..103
 FT /label= as above

PN GB2169295-A.
 PD 09-JUL-1986.

PF 06-JAN-1986; 000199.
 PR 05-OCT-1983; BE-857924.
 PR 07-JAN-1985; GB-000341.
 PR 07-JAN-1985; GB-500341.
 PR 01-NOV-1985; GB-026951.
 PR 06-JAN-1986; GB-000199.
 PA (CELL-) CELTECH LTD.
 PI Harris TJR, Reynolds JJ, Docherty AJP, Murphy G;
 DR WPI; 86-177873/28.
 DR N-PSDB; N60538.
 PT Prod'n. of metallo-proteinase inhibitors - by recombinant DNA
 FT techniques
 PS Disclosure; Fig 3; 16pp; English.
 CC A gene with residues 64-684 of N60538 is claimed. N60538 was
 CC isolated from human a foetal diploid lung cells (ATCC CCL153) cDNA
 CC library using N60539.
 SQ Sequence 207 AA;

Query Match 51.7%; Score 31; DB 1; Length 207;
 Best Local Similarity 50.0%; Pred. No. 62;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
 :|: |: |||
 Db 63 IKTKMKYKGF 72

RESULT 7

ID P60275 standard; Protein; 207 AA.
 AC P60275;
 DT 08-AUG-1991 (first entry)
 DE Sequence of human natural inhibitor of collagenases (NIC).
 KW Metallo-proteinase inhibitor; wound healing; emphysema;
 KW rheumatoid arthritis therapy; ulceration; tumour metastasis.

FH Key Location/Qualifiers
 FT peptide 1..23
 FT /label= signal
 FT protein 24..207
 FT /note= "claimed"
 FT modified_site 53..55
 FT /note= "potential glycosylation site"
 FT modified_site 101..103
 FT /note= "potential glycosylation site"
 PN EP-189784-A.
 PD 06-AUG-1986.
 PF 16-JAN-1986; 100482.
 PR 18-JAN-1985; US-692808.
 PA (SEAR) SEARLE G D & CO.
 PI Galloway WA, Clissold PM, McCullagh KG;
 DR WPI; 86-205910/32.
 DR N-PSDB; N60277.
 PT New human natural inhibitor of collagenase - for treating e.g.
 PT rheumatoid arthritis or ulceration, and new DNA sequences coding
 FT for it

PS Disclosure; Fig 2; 51pp; English.
 CC The patentors claim the AA SQ of human NIC, DNA sequences coding for
 CC NIC, and its RNA analogues and plasmids contg. this DNA. NIC
 CC inhibits the activity of metallo-proteinases, esp. of collagenase,
 CC proteoglycanase, gelatinase or a leucocyte, macrophage or tumour
 CC cell metallo-proteinase.
 SQ Sequence 207 AA;

Query Match 51.7%; Score 31; DB 1; Length 207;
 Best Local Similarity 50.0%; Pred. No. 62;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
 :|: |: |||
 Db 63 IKTKMKYKGF 72

RESULT 8

ID P60592 standard; Protein; 207 AA.
 AC P60592;
 DT 13-AUG-1991 (first entry)
 DE Sequence of a human protein having erythroid potentiating
 DE activity (EPA).
 KW Erythroid precursor growth; anaemia therapy.
 OS Homo sapiens.
 PN WO8602100-A.
 PD 10-APR-1986.
 PF 01-OCT-1985; 501900.
 PR 01-OCT-1984; US-656590.
 PA (SANO) SANDOZ LTD.
 PA (REGC) REGENTS OF UNIV OF CALIF.
 PI Brown E, Clark S, Golde DW, Casson JC, Hewick R, Westbrook CA;
 DR WPI; 86-106663/16.
 DR N-PSDB; N60494.
 PT Vectors contg. gene for protein having erythroid potentiating
 PT activity - used for producing protein to stimulate growth and
 PT formation of erythroid cells
 PS Disclosure; Fig 4; 59pp; English.
 CC The inventors claim human and gibbon EPA proteins (P60592, P60593)
 CC and cDNA that encodes EPA (N60494, N60495). The EPA protein has a
 CC biological activity of at least about 1,000,000 units per mg of
 CC protein and has an apparent molecular weight of about 28,000 daltons.
 SQ Sequence 207 AA;

Query Match 51.7%; Score 31; DB 1; Length 207;
 Best Local Similarity 50.0%; Pred. No. 62;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
 :|: |: |||


```

PR 31-DEC-1990; US-635965.
PA (QUES-) QUEST INT FLAVORS & FOODS INGREDIENTS.
PI Henderson JT, Ledebor AM, Marugg JD, Vanderbergh PA;
DR WPI: 92-227391/28.
DR N-PSDB; Q25810.
PT Cloned gene from Pedococcus acidilactici - encodes bacteriocin
PT used to inhibit Listeria in foods
PS Disclosure; Fig 4; 44pp; English.
CC The protein sequence was deduced from the third ORF encoded by the
CC 5.6 kbp EcoRI-SalI insert from pSR0220. This vector contains three
CC ORFs, the first encoding a bacteriocin precursor, and the others
CC needed for correct protein assembly to obtain functionally active
CC bacteriocin. The bacteriocin from Pedococcus acidilactici NRRL-B-
CC 18050 is used in foods to inhibit bacterial spoilage, e.g. to inhibit
CC Listeria monocytogenes.
CC See also R25810 and R29001.
SQ Sequence 724 AA;

Query Match 51.7%; Score 31; DB 1; Length 724;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 LAKLLKGF 10
   |||||
Db 528 LAKLLVGF 535

RESULT 13
R65005
ID R65005 standard; Protein; 207 AA.
DT 23-OCT-1995 (first entry)
DE Human tissue inhibitor of metalloproteinase (TIMP-1).
KW Tissue inhibitor of metalloproteinase; diagnostic; therapeutic;
OS Homo sapiens.
PN WO9505478-A.
PD 23-FEB-1995.
PF 12-AUG-1994; U09188.
PR 12-AUG-1993; US-105263.
PR 13-DEC-1993; US-167463.
PA (REGC ) UNIV CALIFORNIA.
PI Hawkes SP, Kishnani NS, Yang T;
DR WPI: 95-098775/13.
PT New human tissue inhibitor of metalloproteinase-3 - used to
PT develop prods. for diagnosis, therapy or prophylaxis of
PT conditions with unwanted matrix metalloproteinase activity.
PS Disclosure; Fig 6A-B; 87pp; English.
CC The protein sequence of the human TIMP-1 is compared with those
CC from pig, cattle, rabbit and mouse TIMP-1, human, mouse and cattle
CC TIMP-2 and chicken (Ch) and mouse TIMP-3. A probe based on the
CC CHIMP-3 amino acid sequence (R65000) is used to isolate DNA
CC encoding human TIMP-3 from a human cDNA library. Human TIMP-3
CC can be used for the diagnosis, therapy or prophylaxis of
CC conditions characterized by excess or unwanted matrix
CC metalloproteinase activity, e.g. neoplasias, tumor metastasis,
CC inflammatory disorders such as rheumatoid arthritis, ulcerations,
CC reaction to infection, periodontal disease or osteoporosis. It can
CC also be used in drug screening/design.
SQ Sequence 207 AA;

Query Match 51.7%; Score 31; DB 1; Length 207;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
   ||:|:|
Db 63 IKMTKMYKGF 72

PR 31-DEC-1990; US-635965.
PA (QUES-) QUEST INT FLAVORS & FOODS INGREDIENTS.
PI Henderson JT, Ledebor AM, Marugg JD, Vanderbergh PA;
DR WPI: 92-227391/28.
DR N-PSDB; Q25810.
PT Cloned gene from Pedococcus acidilactici - encodes bacteriocin
PT used to inhibit Listeria in foods
PS Disclosure; Fig 4; 44pp; English.
CC The protein sequence was deduced from the third ORF encoded by the
CC 5.6 kbp EcoRI-SalI insert from pSR0220. This vector contains three
CC ORFs, the first encoding a bacteriocin precursor, and the others
CC needed for correct protein assembly to obtain functionally active
CC bacteriocin. The bacteriocin from Pedococcus acidilactici NRRL-B-
CC 18050 is used in foods to inhibit bacterial spoilage, e.g. to inhibit
CC Listeria monocytogenes.
CC See also R25810 and R29001.
SQ Sequence 724 AA;

Query Match 51.7%; Score 31; DB 1; Length 437;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LAKLLKGF 10
   ||:|:|
Db 121 LAKLLQGF 128

RESULT 15
W20656
ID W20656 standard; protein; 448 AA.
AC W20656;
DE H. pylori transporter protein, 02ge20116orf28.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.
PN WO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI: 97-052306/05.
DR N-PSDB; T67909.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 56; Page 1079-1080; 1481pp; English.

Query Match 51.7%; Score 31; DB 1; Length 437;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LAKLLKGF 10
   ||:|:|
Db 121 LAKLLQGF 128

RESULT 14
W20294
ID W20294 standard; Protein; 437 AA.
AC W20294;
DE H. pylori cell envelope transporter protein 24238762.aa.
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS Helicobacter pylori.
PN WO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI: 97-052306/05.
DR N-PSDB; T67502.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 56; Page 493; 1481pp; English.
CC The present sequence shows a Helicobacter pylori cell envelope
CC transporter protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 437 AA;

```

CC The present sequence is a H. pylori transporter protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 448 AA;

Query Match 51.78; Score 31; DB 1; Length 448;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 3 LAKLLKGF 10
||:|:|:|
Db 132 LARLLQGF 139

Search completed: November 13, 1999, 18:59:57
Job time: 136 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:32 ; Search time 64.87 Seconds
(without alignments)
2.643 Million cell updates/sec

Title: US-08-913-430-10

Perfect score: 60

Sequence: 1 MKLAKLLKGFXXIK 15

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/PTUS9.COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	60.0	8	4	5252328-7
2	31	51.7	207	1	US-08-588-163-5
3	31	51.7	207	1	US-09-111-070-5
4	30	50.0	17	1	US-07-725-331-48
5	30	50.0	17	1	US-07-725-331-49
6	30	50.0	599	1	US-07-879-617A-10
7	30	50.0	599	2	US-08-753-985-10
8	30	50.0	17	3	PCT-US91-05047-48
9	30	50.0	17	3	PCT-US91-05047-49
10	29.5	49.2	22	1	US-07-965-663A-17
11	29	48.3	21	1	US-07-908-455A-13
12	29	48.3	21	1	US-08-434-120-39
13	29	48.3	21	1	US-08-465-325-39
14	29	48.3	731	2	US-08-731-716-2
15	29	48.3	204	2	US-08-808-550-32
16	29	48.3	743	2	US-09-012-030-2
17	29	48.3	413	2	US-08-759-599-12
18	29	48.3	743	2	US-08-590-454-2
19	29	48.3	2254	2	US-08-677-010-3
20	29	48.3	1658	2	US-08-609-049A-13
21	29	48.3	1726	2	US-08-609-049A-30
22	29	48.3	134	2	US-08-847-724-1
23	28	46.7	18	1	US-07-725-331-24
24	28	46.7	17	1	US-07-725-331-51
25	28	46.7	22	1	US-07-725-331-60
26	28	46.7	26	1	US-07-725-331-61
27	28	46.7	30	1	US-07-725-331-62
28	28	46.7	36	1	US-07-725-331-63
29	28	46.7	16	1	US-08-193-521-6
30	28	46.7	16	1	US-07-789-179-2
31	28	46.7	803	1	US-08-158-232-10
32	28	46.7	16	1	US-08-240-712-9
33	28	46.7	803	1	US-08-304-626-10
34	28	46.7	16	1	US-08-434-120-100
35	28	46.7	16	1	US-08-465-325-99
36	28	46.7	16	1	US-08-443-890-9
37	28	46.7	16	1	US-08-444-942-2
38	28	46.7	294	1	US-08-624-125-4
39	28	46.7	803	2	US-08-316-301A-12

Sequence 2, Appli
Sequence 2, Appli
Sequence 10, Appl
Sequence 2, Appli
Sequence 2, Appli
Patent No. 5405775

40 28 46.7 16 2 US-08-446-105-2
41 28 46.7 16 2 US-08-444-939-2
42 28 46.7 803 2 US-08-611-928-10
43 28 46.7 16 2 US-08-444-991-2
44 28 46.7 16 2 US-08-450-733-2
45 28 46.7 100 4 5405775-11

ALIGNMENTS

RESULT 1
5252328-7
; Patent No. 5252328
; APPLICANT: FAULDS, DARYL; VISHOOT, MIMI; BROOKS, EMILY
; TITLE OF INVENTION: MYOPLASMA HYOPNEUMONIAE ANTIGEN AND USES
; THEREFORE
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/335,726
; FILING DATE: 07-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 30,130
; FILING DATE: 26-MAR-1987
; SEQ ID NO: 7:
; LENGTH: 8
5252328-7

Query Match 60.0%; Score 36; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 8.8e-04;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLAKLLK 8
Db 1 MKLAKLLK 8

RESULT 2
US-08-588-163-5
; Sequence 5, Application US/08588163
; Patent No. 5643752
; GENERAL INFORMATION:
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: A NOVEL TISSUE INHIBITOR OF
; TITLE OF INVENTION: METALLOPROTEINASES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,163
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555

TELEFAX: 415-852-0195
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: METALLOPROTEINASES
CLONE: TIMP-1
US-08-588-163-5

Query Match 51.7%; Score 31; DB 1; Length 207;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
:|:|:|
Db 63 IKMTKMYKGF 72

RESULT 3

US-09-111-070-5
; Sequence 5, Application US/09111070
; Patent No. 5914392

GENERAL INFORMATION:

APPLICANT: Hawkins, Phillip R.
APPLICANT: Murry, Lynn E.
TITLE OF INVENTION: A NOVEL TISSUE INHIBITOR OF
TITLE OF INVENTION: METALLOPROTEINASES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/111,070
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/588,163
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33,954
REFERENCE/DOCKET NUMBER: PF-0053
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195

TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: METALLOPROTEINASES
CLONE: TIMP-1
US-09-111-070-5

Query Match 51.7%; Score 31; DB 2; Length 207;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
:|:|:|
Db 63 IKMTKMYKGF 72

RESULT 4

US-07-725-331-48
; Sequence 48, Application US/07725331
; Patent No. 5294605

GENERAL INFORMATION:

APPLICANT: Houghten, Richard
APPLICANT: Blondelle, Sylvie
TITLE OF INVENTION: Amphiphilic Peptide Compositions and
TITLE OF INVENTION: Analogues Thereof
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dressler, Goldsmith, Sutker, Shore,
ADDRESSEE: & Milnamow
STREET: 180 No. 5294605th Stetson
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/725,331
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/554,422
FILING DATE: 19-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: 421250-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3126165418
TELEFAX: 3126165460

SEQUENCE CHARACTERISTICS:

LENGTH: 17 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
OTHER INFORMATION: C-terminal amide, acetylated
OTHER INFORMATION: at N-terminus.

US-07-725-331-48

Query Match 50.0%; Score 30; DB 1; Length 17;
Best Local Similarity 46.7%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKLAKLLKGFXXXIK 15
:|:|:|:|:|
Db 1 LKLKLLKLLKLLK 15

RESULT 5

US-07-725-331-49
; Sequence 49, Application US/07725331
; Patent No. 5294605

GENERAL INFORMATION:
APPLICANT: Houghten, Richard
TITLE OF INVENTION: Amphiphilic Peptide Compositions and Analogues Thereof
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dressier, Goldsmith, Sutker, Shore,
STREET: 180 No. 5294605th Stetson
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/725,331
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA: US 07/554,422
APPLICATION NUMBER: US 07/554,422
FILING DATE: 19-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: 421250-80
TELEPHONE: 3126165418
TELEFAX: 3126165460
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
OTHER INFORMATION: C-terminal amide, acetylated
OTHER INFORMATION: at N-terminus.
US-07-725-331-49

Query Match 50.0%; Score 30; DB 1; Length 17;
Best Local Similarity 46.7%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKLAKLLKGFXXIK 15
: || |||| :
Db 1 LKLLKLLKLLKLLK 15

RESULT 6
US-07-879-617A-10
Sequence 10, Application US/07879617A
Patent No. 5580775
GENERAL INFORMATION:
APPLICANT: Fremean Jr., Robert T.
APPLICANT: Caron, Marc G.
APPLICANT: Blakely, Randy D.
TITLE OF INVENTION: A High Affinity L-Proline Transporter
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30309
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/879,617A
FILING DATE: 19920501
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU109
TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 599 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Rattus
TISSUE TYPE: Brain - GABA transporter
FEATURE:
NAME/KEY: Active-site
LOCATION: 83..84
OTHER INFORMATION: /note= "Leucine zipper motif"
FEATURE:
NAME/KEY: Active-site
LOCATION: 90..91
OTHER INFORMATION: /note= "Leucine zipper motif"
FEATURE:
NAME/KEY: Active-site
LOCATION: 97..98
OTHER INFORMATION: /note= "Leucine zipper motif"
FEATURE:
NAME/KEY: Active-site
LOCATION: 104..105
OTHER INFORMATION: /note= "Leucine zipper motif"
US-07-879-617A-10

Query Match 50.0%; Score 30; DB 1; Length 599;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKG 9
: |||| : ||
Db 116 MKLAPMFKG 124

RESULT 7
US-08-753-985-10
Sequence 10, Application US/08753985
Patent No. 5759788
GENERAL INFORMATION:
APPLICANT: Fremean Jr., Robert T.
APPLICANT: Caron, Marc G.
APPLICANT: Blakely, Randy D.
TITLE OF INVENTION: A High Affinity L-Proline Transporter
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30309

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/753,985
FILING DATE: 03-DEC-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/879617
FILING DATE: 01-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU109
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 599 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Rattus
TISSUE TYPE: Brain - GABA transporter
FEATURE:
NAME/KEY: Active-site
LOCATION: 83..84
OTHER INFORMATION: /note= "Leucine zipper motif"
FEATURE:
NAME/KEY: Active-site
LOCATION: 90..91
OTHER INFORMATION: /note= "Leucine zipper motif"
FEATURE:
NAME/KEY: Active-site
LOCATION: 97..98
OTHER INFORMATION: /note= "Leucine zipper motif"
FEATURE:
NAME/KEY: Active-site
LOCATION: 104..105
OTHER INFORMATION: /note= "Leucine zipper motif"
US-08-753-985-10

Query Match 50.0%; Score 30; DB 2; Length 599;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKG 9
|||| : ||
Db 116 MKLAPMEKG 124

RESULT 8
PCT-US91-05047-48
Sequence 48, Application PC/TUS9105047
GENERAL INFORMATION:
APPLICANT: Houghten, Richard
TITLE OF INVENTION: Amphiphilic Peptide Compositions and
TITLE OF INVENTION: Analogues Thereof
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dressler, Goldsmith, Sutker, Shore,
ADDRESSEE: & Milnamow
STREET: 180 North Stetson

CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/05047
FILING DATE: 19910717
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/554,422
FILING DATE: 19-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gamson, Edward P.
REGISTRATION NUMBER: 29,381
REFERENCE/DOCKET NUMBER: 421250-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3126165418
TELEFAX: 3126165460
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
OTHER INFORMATION: C-terminal amide, acetylated
OTHER INFORMATION: at N-terminus.
PCT-US91-05047-48

Query Match 50.0%; Score 30; DB 3; Length 17;
Best Local Similarity 46.7%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKLAKLLKGXXXIK 15
||| |||| : |
Db 1 LKLLKLLKLLKLLK 15

RESULT 9
PCT-US91-05047-49
Sequence 49, Application PC/TUS9105047
GENERAL INFORMATION:
APPLICANT: Houghten, Richard
TITLE OF INVENTION: Amphiphilic Peptide Compositions and
TITLE OF INVENTION: Analogues Thereof
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dressler, Goldsmith, Sutker, Shore,
ADDRESSEE: & Milnamow
STREET: 180 North Stetson
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/05047
FILING DATE: 19910717
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/554,422

;; FILING DATE: 19-JUL-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gamson, Edward P.
;; REGISTRATION NUMBER: 29,381
;; REFERENCE/DOCKET NUMBER: 421250-80
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 3126165418
;; TELEFAX: 3126165460
;; INFORMATION FOR SEQ ID NO: 49:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 17 amino acids
;; TYPE: AMINO ACID
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; FEATURE:
;; OTHER INFORMATION: C-terminal amide, acetylated
;; OTHER INFORMATION: at N-terminus.
PCT-US91-05047-49

Query Match 50.0%; Score 30; DB 3; Length 17;
Best Local Similarity 46.7%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKLAKLLKGFXXIK 15
:||| ||| :|
Db 1 LKLLKLLKLLKLLK 15

RESULT 10
US-07-965-663A-17
; Sequence 17, Application US/07965663A
; Patent No. 5424290
; GENERAL INFORMATION:
; APPLICANT: Lee, Maloy W.
; APPLICANT: Prasad, Karl U.
; TITLE OF INVENTION: No. 5424290el Biologically Active Peptides and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/965,663A
; FILING DATE: 26-OCT-1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0039-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: /note= "May be a C-terminal amide, and/or may

US-07-965-663A-17

Query Match 49.2%; Score 29.5; DB 1; Length 22;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 8; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 1 MKLA-KLLKGFXXIK 15
:||| | ||| :|
Db 6 LKLAKRFKGFKKLK 21

RESULT 11
US-07-908-455A-13
; Sequence 13, Application US/07908455A
; Patent No. 5459237
; GENERAL INFORMATION:
; APPLICANT: Berkowitz, Barry A.
; APPLICANT: Karl, U. Prasad
; APPLICANT: Maloy, W. Lee
; TITLE OF INVENTION: No. 5459237el Peptide Compositions and
; TITLE OF INVENTION: Uses Therefor
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi & Stewart
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: DW4.V2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/908,455A
; FILING DATE: 19920702
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07686115
; FILING DATE: 15-APR-1991
; APPLICATION NUMBER: US 07476629
; FILING DATE: 08-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 421250-122
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: amide-terminated
US-07-908-455A-13

Query Match 48.3%; Score 29; DB 1; Length 21;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 AKLLKGFXXIK 15
||: ||| |
Db 3 AKIAGFAKIAK 14

RESULT 12
US-08-434-120-39
; Sequence 39, Application US/08434120
; Patent No. 5635479
; GENERAL INFORMATION:
; APPLICANT: Baker, Margaret A.
; APPLICANT: Jacob, Leonard S.
; APPLICANT: Maloy, W. Lee
; TITLE OF INVENTION: Treatment of Gynecological
; TITLE OF INVENTION: Malignancies with
; TITLE OF INVENTION: Biologically Active Peptides
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: DM4.V2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,120
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,950
; FILING DATE:
; APPLICATION NUMBER: US/08/226,108
; FILING DATE:
; APPLICATION NUMBER: US/07/937,462
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 421250-194
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-434-120-39

Query Match 48.3%; Score 29; DB 1; Length 21;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 AKLLKGFXKXIK 15
||: ||| |
Db 3 AKIAGFKAKI 14

RESULT 13
US-08-465-325-39
; Sequence 39, Application US/08465325
; Patent No. 5686563
; GENERAL INFORMATION:
; APPLICANT: Magainin Pharmaceuticals Inc.
; APPLICANT: 5110 Campus Drive
; APPLICANT: Plymouth Meeting, PA 19462
; TITLE OF INVENTION: Biologically Active Peptides Having
; TITLE OF INVENTION: N-Terminal Substitutions
; NUMBER OF SEQUENCES: 153

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I. Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,325
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/184,462
; FILING DATE: 18-JAN-94
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/891,201
; FILING DATE: 01-JUN-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0021-03000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-465-325-39

Query Match 48.3%; Score 29; DB 1; Length 21;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 AKLLKGFXKXIK 15
||: ||| |
Db 3 AKIAGFKAKI 14

RESULT 14
US-08-731-716-2
; Sequence 2, Application US/08731716
; Patent No. 5789202
; GENERAL INFORMATION:
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Jaskunas, S. Richard
; APPLICANT: Rockey, Pamela K.
; APPLICANT: Zhao, Genshi
; APPLICANT: Rostock, Paul R. Jr.
; APPLICANT: No. 5789202ris, Franklin H.
; TITLE OF INVENTION: Penicillin Binding Protein From
; TITLE OF INVENTION: Streptococcus Pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/731,716
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Webster, Thomas D.
 REGISTRATION NUMBER: 39,872
 REFERENCE/DOCKET NUMBER: X-10,887
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-276-3334
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 731 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-731-716-2

Query Match 48.3%; Score 29; DB 2; Length 731;
 Best Local Similarity 46.7%; Pred. No. 4.2e+02;
 Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 MKLAKLKGFXXIK 15
 ||| || | |
 DB 1 MKLDKLFKFLSLFK 15

RESULT 15
 US-08-808-550-32
 Sequence 32, Application US/080808550
 Patent No. 5871992
 GENERAL INFORMATION:
 APPLICANT: Teebor, George W.
 APPLICANT: Hilbert, Timothy P.
 TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND
 TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
 NUMBER OF SEQUENCES: 42
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: David A. Jackson, Esq.
 STREET: 411 Hackensack Ave, Continental Plaza, 4th
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/808,550
 FILING DATE: 26-FEB-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 1049-1-001 N
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 204 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 FRAGMENT TYPE: N-terminal
 ORIGINAL SOURCE:

ORGANISM: M. jannaschii
 US-08-808-550-32

Query Match 48.3%; Score 29; DB 2; Length 204;
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLAKLIK 8
 ||||| |
 DB 87 KLAKILK 93

Search completed: November 13, 1999, 10:56:33
 Job time: 1362 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:38 ; Search time 251.81 Seconds
(without alignments)
3.771 Million cell updates/sec

Title: US-08-913-430-10

Perfect score: 60
Sequence: 1 MKLAKLLKGFXXIK 15

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA:*
1: /cgn2_6/ptodata/1/paa/PCTUS99_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	90.0	15	13	US-08-913-430-10
2	36	60.0	904	1	PCT-US98-25044-2
3	36	60.0	361	1	PCT-US98-25044-3
4	36	60.0	191	1	PCT-US98-25044-5
5	36	60.0	205	4	US-08-134-231A-25
6	36	60.0	205	4	US-08-167-463A-8
7	36	60.0	205	4	US-08-167-463A-8
8	36	60.0	205	11	US-08-728-160-25
9	36	60.0	339	15	US-09-107-532-7148
10	36	60.0	361	15	US-09-198-484-2
11	36	60.0	361	15	US-09-198-484-3
12	36	60.0	191	15	US-09-198-484-5
13	33	55.0	77	16	US-09-248-796-26324
14	33	55.0	77	19	US-09-096-409-26324
15	32	53.3	207	4	US-08-134-231A-22
16	32	53.3	207	4	US-08-167-463A-4
17	32	53.3	207	4	US-08-167-463A-5
18	32	53.3	207	4	US-08-167-463A-4
19	32	53.3	207	4	US-08-167-463A-5

20	53.3	207	11	US-08-728-160-22	Sequence 22, Appl
21	53.3	133	11	US-08-761-066-421	Sequence 421, Appl
22	53.3	375	15	US-09-107-532-5400	Sequence 5400, Ap
23	53.3	388	15	US-09-134-000-4774	Sequence 4774, Ap
24	53.3	283	16	US-09-248-796-14990	Sequence 14990, A
25	53.3	153	16	US-09-270-767-32945	Sequence 32945, A
26	53.3	153	16	US-09-270-767-48162	Sequence 48162, A
27	53.3	193	16	US-09-299-708-56	Sequence 56, Appl
28	53.3	283	19	US-60-096-409-14990	Sequence 14990, A
29	53.3	111	19	US-60-140-806-864	Sequence 864, App
30	53.3	228	19	US-60-143-868-523	Sequence 523, App
31	53.3	676	23	US-09-417-507-37196	Sequence 37196, A
32	51.7	207	4	US-08-134-231A-23	Sequence 23, Appl
33	51.7	207	4	US-08-167-463A-6	Sequence 6, Appli
34	51.7	207	4	US-08-167-463A-6	Sequence 6, Appli
35	51.7	207	8	US-08-463-261A-11	Sequence 11, Appl
36	51.7	437	8	US-08-487-032A-593	Sequence 593, App
37	51.7	437	9	US-08-561-469A-593	Sequence 593, App
38	51.7	207	11	US-08-728-160-23	Sequence 23, Appl
39	51.7	437	11	US-08-761-184-966	Sequence 966, App
40	51.7	448	11	US-08-761-184-1115	Sequence 1115, Ap
41	51.7	437	11	US-08-761-184-1380	Sequence 1380, Ap
42	51.7	437	12	US-08-821-931-966	Sequence 966, App
43	51.7	448	12	US-08-821-931-1115	Sequence 1115, Ap
44	51.7	437	12	US-08-821-931-1380	Sequence 1380, Ap
45	51.7	505	16	US-09-270-767-43185	Sequence 43185, A

ALIGNMENTS

RESULT 1
US-08-913-430-10
; Sequence 10, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0 - beta
; SEQ ID NO 10
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: (11)
; OTHER INFORMATION: Undetermined
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (12)
; OTHER INFORMATION: Residue may be Asn or Leu
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (13)
; OTHER INFORMATION: Residue may be Met or Val
; US-08-913-430-10

Query Match 90.0%; Score 54; DB 13; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00039;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLAKLLKGFXXIK -15
|||||

Db 1 MKLAKLLKGFXXXIK 15

RESULT 2
PCT-US98-25044-2
; Sequence 2, Application PC/TUS9825044
; GENERAL INFORMATION:
; APPLICANT: Iowa State University Research Foundation
; TITLE OF INVENTION: RECOMBINANT MYCOPLASMA HYOPNEUMONIAE VACCINE
; FILE REFERENCE: 19000.028/P028-PCT
; CURRENT APPLICATION NUMBER: PCT/US98/25044
; CURRENT FILING DATE: 1998-11-24
; .NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 904
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
PCT-US98-25044-2

Query Match 60.0%; Score 36; DB 1; Length 904;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLAKLLK 8
| | | | | | | |
Db 1 MKLAKLLK 8

RESULT 3
PCT-US98-25044-3
; Sequence 3, Application PC/TUS9825044
; GENERAL INFORMATION:
; APPLICANT: Iowa State University Research Foundation
; TITLE OF INVENTION: RECOMBINANT MYCOPLASMA HYOPNEUMONIAE VACCINE
; FILE REFERENCE: 19000.028/P028-PCT
; CURRENT APPLICATION NUMBER: PCT/US98/25044
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 361
; TYPE: PRT
; ORGANISM: protein encoded by a clone of Mycoplasma hyopneumoniae
PCT-US98-25044-3

Query Match 60.0%; Score 36; DB 1; Length 361;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLAKLLK 8
| | | | | | | |
Db 1 MKLAKLLK 8

RESULT 4
PCT-US98-25044-5
; Sequence 5, Application PC/TUS9825044
; GENERAL INFORMATION:
; APPLICANT: Iowa State University Research Foundation
; TITLE OF INVENTION: RECOMBINANT MYCOPLASMA HYOPNEUMONIAE VACCINE
; FILE REFERENCE: 19000.028/P028-PCT
; CURRENT APPLICATION NUMBER: PCT/US98/25044
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 191
; TYPE: PRT
; ORGANISM: protein encoded by a clone of Mycoplasma hyopneumoniae
PCT-US98-25044-5

Query Match 60.0%; Score 36; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLAKLLK 8
| | | | | | | |
Db 1 MKLAKLLK 8

RESULT 5
US-08-134-231A-25
; Sequence 25, Application US/08134231A
; GENERAL INFORMATION:
; APPLICANT: Silbiger, Scott M.
; APPLICANT: KOSKI, Raymond A.
; TITLE OF INVENTION: Tissue Inhibitor Metalloproteinase Type
; TITLE OF INVENTION: Three (TIMP-3) Composition and Methods
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Angen Inc./Patent Operations/KMP
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/134,231A
; FILING DATE:
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-134-231A-25

Query Match 60.0%; Score 36; DB 4; Length 205;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLAKLLKGF 10
| | | | | | | |
Db 64 IKMTKMLKGF 73

RESULT 6
US-08-167-463-8
; Sequence 8, Application US/08167463
; GENERAL INFORMATION:
; APPLICANT: HAWKES, SUSAN P.
; APPLICANT: KISHNANI, NARENDRA S.
; APPLICANT: YANG, TE-TUAN
; TITLE OF INVENTION: HUMAN TIMP-3
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/167,463
; FILING DATE: 13-DEC-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: LEHNHARDT, SUSAN K.
; REGISTRATION NUMBER: 33,943
; REFERENCE/DOCKET NUMBER: 22000-20542.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-167-463-8

Query Match 60.08; Score 36; DB 4; Length 205;
Best Local Similarity 60.08; Pred. No. 26;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
Db 64 IKMTKMLKGF 73

RESULT 7
US-08-167-463A-8
; Sequence 8, Application US/08167463A
; GENERAL INFORMATION:
; APPLICANT: HAWKES, SUSAN P.
; APPLICANT: KISHNANI, NARENDRA S.
; APPLICANT: YANG, TE-TUAN
; TITLE OF INVENTION: HUMAN TIMP-3
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/167,463A
; FILING DATE: 13-DEC-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: LEHNHARDT, SUSAN K.
; REGISTRATION NUMBER: 33,943
; REFERENCE/DOCKET NUMBER: 22000-20542.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-167-463A-8
```

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/167,463
; FILING DATE: 13-DEC-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: LEHNHARDT, SUSAN K.
; REGISTRATION NUMBER: 33,943
; REFERENCE/DOCKET NUMBER: 22000-20542.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-167-463-8

Query Match 60.0%; Score 36; DB 4; Length 205;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
Db 64 IKMTKMLKGF 73

RESULT 8
US-08-728-160-25
; Sequence 25, Application US/08728160
; GENERAL INFORMATION:
; APPLICANT: Silbiger, Scott M.
; APPLICANT: Koski, Raymond A.
; TITLE OF INVENTION: Tissue Inhibitor Metalloproteinase Type
; TITLE OF INVENTION: Three (TIMP-3) Composition and Methods
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc./Patent Operations/KMP
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,160
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/134,231
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-728-160-25

Query Match 60.0%; Score 36; DB 11; Length 205;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
Db 64 IKMTKMLKGF 73

RESULT 9
US-09-107-532-7148
; Sequence 7148, Application US/09107532
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7308
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
```

ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER:
OPERATING SYSTEM:
SOFTWARE:
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107.532
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085598
FILING DATE: May 14, 1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Daneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 7148:
SEQUENCE CHARACTERISTICS:
LENGTH: 339 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...339
US-09-107-532-7148

Query Match 60.0%; Score 36; DB 15; Length 339;
Best Local Similarity 61.5%; Pred. No. 46;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 KLAKLLGFXXXI 14
:|||||
Db 50 ELAELLKGFNIII 62

RESULT 10
US-09-198-484-2
; Sequence 2, Application US/09198484
; GENERAL INFORMATION:
; APPLICANT: Minion, F. Chris
; TITLE OF INVENTION: RECOMBINANT MYCOPLASMA HYOPNEUMONIAE VACCINE
; FILE REFERENCE: 19000.028/P028
; CURRENT APPLICATION NUMBER: US/09/198.484
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 904
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
US-09-198-484-2

Query Match 60.08%; Score 36; DB 15; Length 904;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLAKLLK 8
:|||||
Db 1 MKLAKLLK 8

RESULT 11
US-09-198-484-3
; Sequence 3, Application US/09198484
; GENERAL INFORMATION:
; APPLICANT: Minion, F. Chris
; TITLE OF INVENTION: RECOMBINANT MYCOPLASMA HYOPNEUMONIAE VACCINE
; FILE REFERENCE: 19000.028/P028
; CURRENT APPLICATION NUMBER: US/09/198.484
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 361
; TYPE: PRT
; ORGANISM: protein encoded by a clone of Mycoplasma hyopneumoniae
US-09-198-484-3

Query Match 60.0%; Score 36; DB 15; Length 361;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLAKLLK 8
:|||||
Db 1 MKLAKLLK 8

RESULT 12
US-09-198-484-5
; Sequence 5, Application US/09198484
; GENERAL INFORMATION:
; APPLICANT: Minion, F. Chris
; TITLE OF INVENTION: RECOMBINANT MYCOPLASMA HYOPNEUMONIAE VACCINE
; FILE REFERENCE: 19000.028/P028
; CURRENT APPLICATION NUMBER: US/09/198.484
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 191
; TYPE: PRT
; ORGANISM: protein encoded by a clone of Mycoplasma hyopneumoniae
US-09-198-484-5

Query Match 60.0%; Score 36; DB 15; Length 191;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLAKLLK 8
:|||||
Db 1 MKLAKLLK 8

RESULT 13
US-09-248-796-26324
; Sequence 26324, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 26324
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-26324

Query Match 53.3%; Score 32; DB 4; Length 207;
Best Local Similarity 50.0%; Pred. NO. 1.6e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
Db 63 IKWTRMKFGF 72

Search completed: November 13, 1999, 05:08:38
Job time: 10864 sec

Query Match 55.0%; Score 33; DB 16; Length 77;
Best Local Similarity 55.6%; Pred. NO. 31;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLAKLLKGF 10
Db 1 KVSMMKGF 9

RESULT 14
US-60-096-409-26324
; Sequence 26324, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 26324
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-26324

Query Match 55.0%; Score 33; DB 19; Length 77;
Best Local Similarity 55.6%; Pred. NO. 31;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLAKLLKGF 10
Db 1 KVSMMKGF 9

RESULT 15
US-08-134-231A-22
; Sequence 22, Application US/08134231A
; GENERAL INFORMATION:
; APPLICANT: Silbiger, Scott M.
; APPLICANT: Koski, Raymond A.
; TITLE OF INVENTION: Tissue Inhibitor Metalloproteinase Type
; TITLE OF INVENTION: Three (TIMP-3) Composition and Methods
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc./Patent Operations/KMP
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/134,231A
; FILING DATE:
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-134-231A-22

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:08:03 ; Search time 75.45 Seconds
(without alignments)
7.965 Million cell updates/sec

Title: US-08-913-430-10
Perfect score: 60
Sequence: 1 MKLAKLLKGFXXIX 15

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR-60.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	37	61.7	2166	2 G70163	hypothetical prote
2	36	60.0	205	1 A26106	metalloproteinase
3	36	60.0	217	2 JC2557	metalloproteinase
4	35	58.3	109	2 B64492	hypothetical prote
5	35	58.3	1022	2 S50534	hypothetical prote
6	35	58.3	173	2 E71017	probable NADH-ubiq
7	34	56.7	913	2 S61580	probable membrane
8	33	55.0	89	1 CCS86	cytochrome c6 - Sp
9	33	55.0	1966	1 MKRW	myosin heavy chain
10	33	55.0	90	2 I76722	phosphocarrier pro
11	33	55.0	255	2 S30999	gene 54 protein -
12	33	55.0	481	2 S47091	cyclase-associated
13	33	55.0	436	2 I55465	[pyruvate dehydrog
14	33	55.0	434	2 A49886	[pyruvate dehydrog
15	33	55.0	440	2 B71162	hypothetical prote
16	32	53.3	26926	1 I38344	titin, cardiac mus
17	32	53.3	207	1 A35685	metalloproteinase
18	32	53.3	6805	2 S20901	titin - rabbit (fr
19	32	53.3	1161	2 B70387	DNA polymerase III
20	32	53.3	207	2 I45964	TIMP-1 - sheep
21	32	53.3	207	2 I47061	collagenase inhibi
22	32	53.3	605	2 E71314	probable GTP-bindi
23	32	53.3	269	2 G69967	hemolysin homolog
24	32	53.3	379	2 S09791	hypothetical prote
25	32	53.3	133	2 D64609	conserved hypothet
26	32	53.3	133	2 H71905	hypothetical prote
27	32	53.3	383	2 G71648	alanine dehydrogen
28	32	53.3	322	2 S18999	asparaginase (EC 3
29	32	53.3	266	2 E69746	conserved hypothet
30	32	53.3	476	2 A56510	dolichyl-diphospho
31	32	53.3	837	2 S64785	hypothetical prote
32	32	53.3	312	2 S67052	hypothetical prote
33	32	53.3	579	2 S69069	hypothetical prote
34	32	53.3	2496	2 A71616	secreted protein p
35	32	53.3	907	2 A45560	sporozoite surface
36	32	53.3	361	2 C71242	hypothetical prote
37	31	51.7	752	1 P11VBL	RNA-directed RNA p
38	31	51.7	752	1 P11VBC	RNA-directed RNA p
39	31	51.7	752	1 P11VBW	RNA-directed RNA p

40 31 51.7 207 1 ZYHUEP metalloproteinase
41 31 51.7 426 1 B71249 hypothetical prote
42 31 51.7 483 2 S61975 glutathione reduct
43 31 51.7 458 2 I39477 glutathione reduct
44 31 51.7 743 2 D64062 GTP pyrophosphokin
45 31 51.7 184 2 S00957 hypothetical prote

ALIGNMENTS

RESULT 1
G70163
Hypothetical protein BB0512 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
C:Accession: G70163
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: G70163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2166 <KLE>
A:Cross-references: GB:AE001153; GB:AE000783; NID:g2688419; PID:g2688426; TIGR:BB0512
A:Experimental source: strain B31

Query Match 61.7%; Score 37; DB 2; Length 2166;
Best Local Similarity 61.5%; Pred No. 45;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 KLAKLLKGFXXIX 14
DB 131 KLSKTLKGFNTQI 143
RESULT 2
A26106
metalloproteinase inhibitor 1 precursor - mouse
N:Alternate names: erythroid potentiating activity (EPA); fibroblast collagenase inhi
C:Species: Mus musculus (house mouse)
C>Date: 05-Oct-1988 #sequence_revision 12-Apr-1996 #text_change 20-Mar-1998
C:Accession: A26917; A26106; A26633; A05276
R:Johnson, M.D.; Housey, G.M.; Kirschmeier, P.T.; Weinstein, I.B.
Mol. Cell. Biol. 7, 2821-2829, 1987
A:Title: Molecular cloning of gene sequences regulated by tumor promoters and mitogen
A:Reference number: A26917; MUID:88038821
A:Accession: A26917
A:Molecule type: mRNA
A:Residues: 1-205 <JOH>
A:Cross-references: GB:M17243; NID:g202111; PID:g202112
A:Experimental source: embryonic fibroblast cDNA library
R:Edwards, D.R.; Waterhouse, P.; Holman, M.L.; Denhardt, D.T.
Nucleic Acids Res. 14, 8863-8878, 1986
A:Title: A growth-responsive gene (16C8) in normal mouse fibroblasts homologous to a
pts.

A:Reference number: A26106; MUID:87066763
A:Accession: A26106
A:Molecule type: mRNA
A:Residues: 1-205 <EDW>
A:Cross-references: GB:X04684; NID:g49702; PID:g49704
A:Experimental source: embryonic fibroblast cDNA library
R:Gewert, D.R.; Coulombe, B.; Castellino, M.; Skup, D.; Williams, B.R.G.
EMBO J. 6, 651-657, 1987
A:Title: Characterization and expression of a murine gene homologous to human EPA/TIM
A:Reference number: A26633; MUID:87218524
A:Accession: A26633
A:Molecule type: DNA

A:Residues: 1-51,'R',53-66,'M',67-116,'KF',119-120,'N',122-138,'V',140-142,'KN',144-193,
 A:Cross-references: GB:M28312; NID:g193040; PID:g193042
 R:Skup, D.; Windass, J.D.; Sor, F.; George, H.; Williams, B.R.G.; Fukuhara, H.; De Maeyer
 Nucleic Acids Res. 10, 3069-3084, 1982
 A:Title: Molecular cloning of partial cDNA copies of two distinct mouse IFN-beta mRNAs.
 A:Reference number: A93424; MUID:82247191

A:Accession: A05276
 A:Molecule type: mRNA
 A:Residues: 158-193,'L',195-205 <SKU>
 A:Note: authors thought this clone represented a form of interferon beta
 C:Genetics:

A:Gene: Timp
 A:Map position: X
 A:Introns: 41/3; 68/3; 110/3; 152/3
 C:Function:

A:Description: regulation of extracellular matrix remodeling by inhibition of matrix met
 Possibly controlling their activation; TIMP-1 and TIMP-2 possess erythroid potentiating
 A:Note: transcription induced by cytokines, tumor promoters, and anti-inflammatory agent
 A:Note: TIMP-1 and TIMP-3 have distinct but overlapping tissue-specific expression patte
 C:Superfamily: metalloproteinase inhibitor

C:Keywords: erythropoiesis; extracellular matrix; glycoprotein; metalloproteinase inhibi
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-205/Product: metalloproteinase inhibitor 1 #status predicted <MAT>
 F:25-94,27-123,37-148,151-197,156-161,169-189/disulfide bonds: #status predicted
 F:54,102/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 60.0%; Score 36; DB 1; Length 205;
 Best Local Similarity 60.0%; Pred. No. 6.4;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
 :|:|:|:|:|
 DB 64 IKMTKMLKGF 73

RESULT 3
 JC2557
 metalloproteinase inhibitor 1 precursor - rat
 N:Alternate names: TIMP-1
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 16-May-1995 #sequence_revision 01-Dec-1995 #text_change 08-Sep-1997
 C:Accession: JC2557; B39120; S20326
 R:Okada, A.; Garnier, J.M.; Vicaire, S.; Basset, P.
 Gene 147, 301-302, 1994
 A:Title: Cloning of the cDNA encoding rat tissue inhibitor of metalloproteinase 1 (TIMP-
 A:Reference number: JC2557; MUID:95011636
 A:Accession: JC2557
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-217 <OKA>
 A:Cross-references: EMBL:U06179; NID:g468057; PID:g468058
 B:Olson Jr., J.A.; Shiverick, K.T.; Ogilvie, S.; Buhl, W.C.; Raizada, M.K.
 Proc. Natl. Acad. Sci. U.S.A. 88, 1928-1932, 1991
 A:Title: Angiotensin II induces secretion of plasminogen activator inhibitor 1 and a tis
 A:Reference number: B39120; MUID:91156719
 A:Accession: B39120
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 24-36,'B',38,'B',40-42,'X',44 <OLS>
 R:Roswit, W.T.; McCourt, D.W.; Partridge, N.C.; Jeffrey, J.J.
 Arch. Biochem. Biophys. 292, 402-410, 1992
 A:Title: Purification and sequence analysis of two rat tissue inhibitors of metalloprote
 A:Reference number: S20325; MUID:92117648
 A:Accession: S20326
 A:Molecule type: protein
 A:Residues: 24-45 <ROS>
 C:Function:

A:Description: regulation of extracellular matrix remodeling by inhibition of matrix met
 Possibly controlling their activation; TIMP-1 and TIMP-2 possess erythroid potentiating
 A:Note: transcription induced by cytokines, tumor promoters, and anti-inflammatory agent
 A:Note: TIMP-1 and TIMP-3 have distinct but overlapping tissue-specific expression patte
 C:Superfamily: metalloproteinase inhibitor

C:Keywords: erythropoiesis; extracellular matrix; glycoprotein; metalloproteinase inh
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-217/Product: metalloproteinase inhibitor 1 #status experimental <MAT>
 F:101,130/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 60.0%; Score 36; DB 2; Length 217;
 Best Local Similarity 60.0%; Pred. No. 6.8;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
 :|:|:|:|:|
 DB 63 IKMTKMLKGF 72

RESULT 4
 B64492
 hypothetical protein M01539 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997
 C:Accession: B64492
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
 ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
 A:Reference number: A64300; MUID:96337999
 A:Accession: B64492
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-109 <BUL>
 A:Cross-references: GB:U67595; GB:L77117; NID:g1592170; PID:g1592171; TIGR:M01539; PI
 C:Genetics:

A:Map position: FOR1516718-1517047
 A:Start codon: GTG

Query Match 58.3%; Score 35; DB 2; Length 109;
 Best Local Similarity 50.0%; Pred. No. 5.3;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 KLAKLLKGFXXXIK 15
 :|:|:|:|:|
 DB 71 KLAKILGGSGSEVK 84

RESULT 5
 S50534
 hypothetical protein YEL055c - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 06-Feb-1998
 C:Accession: S50534
 R:Dietrich, F.S.
 submitted to the EMBL Data Library, December 1994
 A:Description: The sequence of S. cerevisiae cosmid 9669, 8334, 8199, and lambda clo

A:Reference number: S50428
 A:Accession: S50534
 A:Molecule type: DNA
 A:Residues: 1-1022 <DIE>
 A:Cross-references: EMBL:U18795; NID:g603241; PID:g603263; MIPS:YEL055c
 C:Genetics:
 A:Gene: SGD:POL5
 A:Cross-references: SGD:S0000781; MIPS:YEL055c
 A:Map position: 5L

Query Match 58.3%; Score 35; DB 2; Length 1022;
 Best Local Similarity 57.1%; Pred. No. 51;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 KLAKLLKGFXXXIK 15
 :|:|:|:|:|

Db 906 KISKLLKGKIFKIK 919

RESULT 6

E71017
probable NADH-ubiquinone oxidoreductase subunit - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: E71017
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ofuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5: 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137
A:Accession: E71017
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-173 <KAW>
A:Cross-references: GB:AP000006; NID:93236133; PID:d1031484; PID:g3257858
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1434

Query Match 58.3%; Score 35; DB 2; Length 173;

Best Local Similarity 87.5%; Pred. No. 8.4;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLAKLLKG 9

|||||

Db 157 KLAKMLKG 164

RESULT 7

S61580
probable membrane protein YDR213w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YD8142B.05
C:Species: Saccharomyces cerevisiae
C>Date: 09-Mar-1996 #sequence_revision 12-Apr-1996 #text_change 12-Dec-1997
C:Accession: S61580
R:Oliver, K.; Harris, D.
Submitted to the EMBL Data Library, December 1995
A:Reference number: S61576
A:Accession: S61580
A:Molecule type: DNA
A:Residues: 1-913 <OLI>
A:Cross-references: EMBL:268195; NID:g1122341; PID:e213844; PID:g1122346; MIPS:YDR213w
A:Experimental source: strain AB972
C:Genetics:
A:Map position: 4R
C:Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluster h
C:Keywords: transmembrane protein
F:46-83/Domain: GAL4 zinc binuclear cluster homology <GAL4>
F:697-713/Domain: transmembrane #status predicted <TM>

Query Match 56.7%; Score 34; DB 2; Length 913;

Best Local Similarity 54.5%; Pred. No. 72;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 KLLKGFXKXIK 15

|||||

Db 853 KLLRGFAFEVK 863

RESULT 8

CCSG6
cytochrome c6 - Spirulina maxima
N:Alternate names: cytochrome c553; soluble cytochrome f
C:Species: Spirulina maxima
C>Date: 07-May-1991 #sequence_revision 07-May-1981 #text_change 15-Jan-1999
C:Accession: A00110

R:Ambler, R.P.; Bartsch, R.G.
Nature 253, 285-288, 1975
A:Title: Amino acid sequence similarity between cytochrome f from a blue-green bacter
A:Reference number: A00110; MUID:75100362
A:Accession: A00110
A:Molecule type: protein
A:Residues: 1-89 <AMB>
C:Superfamily: cytochrome c6; cytochrome c6 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; photosynthesis
F:4-81/Domain: cytochrome c6 homology <CYC>
F:14,17/Binding site: heme (Cys) (covalent) #status predicted
F:18,62/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 55.0%; Score 33; DB 1; Length 89;

Best Local Similarity 87.5%; Pred. No. 11;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 LAKLLKGF 10

|||||

Db 36 LAKYLKGF 43

RESULT 9

MWKW
myosin heavy chain B - Caenorhabditis elegans
N:Contains: myosin ATPase (EC 3.6.1.32)
C:Species: Caenorhabditis elegans
C>Date: 13-Jun-1983 #sequence_revision 19-Feb-1984 #text_change 20-Mar-1998
C:Accession: A93958; A93287; A21074; A02992
R:Karn, J.; Brenner, S.; Barnett, L.
Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983
A:Title: Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy
A:Reference number: A93958; MUID:83273600
A:Accession: A93958
A:Molecule type: DNA
A:Residues: 1-1966 <KAR>
A:Cross-references: GB:J01050; NID:g156399; PID:gi56400
R:McLachlan, A.D.; Karn, J.
Nature 299, 226-231, 1982
A:Title: Periodic charge distributions in the myosin rod amino acid sequence match cr
A:Reference number: A93287; MUID:82272395
A:Accession: A93287
A:Molecule type: DNA
A:Residues: 850-1336, 'R', 1338-1879, 'L', 1881-1966 <MCI>
R:Wills, N.; Gesteland, R.F.; Karn, J.; Barnett, L.; Bolten, S.; Waterston, R.H.
Cell 33, 575-583, 1983
A:Title: The genes sup-7 x and sup-5 III of Caenorhabditis elegans suppress amber non
A:Reference number: A21074; MUID:83232892
A:Accession: A21074
A:Molecule type: DNA
A:Residues: 1876-1966 <WIL>
A:Cross-references: GB:V01494; GB:J01049; NID:g6783; PID:g6784
C:Genetics:
A:Gene: unc-54
A:Introns: 21/3; 64/2; 114/3; 267/1; 528/3; 1750/3; 1822/3; 1897/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle
F:87-778/Domain: myosin motor domain homology <MMOT>
F:177-184/Region: nucleotide-binding motif A (P-loop)
F:665-687/Region: actin binding #status predicted
F:769-783/Region: actin binding #status predicted
F:851-1966/Domain: actin binding #status predicted
F:851-1165/Region: S2
F:1166-1966/Region: light meromyosin
F:128/Modified site: N6, N6-trimethyllysine (Lys) #status predicted
F:183/Binding site: ATP (Lys) #status predicted
F:705,715/Active site: Cys #status predicted

Query Match 55.0%; Score 33; DB 1; Length 1966;

Best Local Similarity 50.0%; Pred. No. 2.5e+02;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;


```
Qy 2 KLAALLKGFXXXIK 15
    ||| :| :| :| :
Db 792 KLAITLGFQSQIR 805

RESULT 10
176722
Phosphocarrier protein Npr (nitrogen related hpr) - Escherichia coli
C:Species: Escherichia coli
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 20-Feb-1998
C:Accession: I76722; H65111; S38619
R:Jones, D.H.A.; Franklin, C.F.H.; Thomas, C.M.
Microbiology 140, 1035-1043, 1994
A:Title: Molecular analysis of the operon which encodes the RNA polymerase sigma factor
A:Reference number: I57054; MUID:94297724
A:Accession: I76722
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-90 <RES>
A:Cross-references: EMBL:Z27094; NID:G414884; PID:G414889
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H65111
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-90 <BLAT>
A:Cross-references: GB:AF000400; GB:U00096; NID:G2367203; PID:G1789599; UWGP:B3206
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ptsO
C:Superfamily: phosphotransferase system phosphohistidine-containing protein; phosphotran
F:9-86/Domain: phosphotransferase system phosphohistidine-containing protein homology <R
F:16/Binding site: phosphate (His) (covalent) (by phosphotransferase system enzyme I) #s
F:48/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 55.0%; Score 33; DB 2; Length 90;
Best Local Similarity 42.9%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MKLAKLLKGFXXXI 14
    ||| :| :| :| :
Db 21 MKLFELMQGFDAEV 34

RESULT 11
S30959
gene 54 protein - Mycobacterium phage L5
C:Species: Mycobacterium phage L5
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 29-May-1998
C:Accession: S30959
R:Donnelly-Wu, M.K.; Jacobs Jr., W.R.; Hatfull, G.F.
Mol. Microbiol. 7, 407-417, 1993
A:Title: Superinfection immunity of mycobacteriophage L5: applications for genetic trans
A:Reference number: S30949
A:Accession: S30959
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-255 <DON>
A:Cross-references: EMBL:Z18946; NID:G15859; PID:e59672; PID:G579137
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, December 1992
C:Comment: In this sequence the first or second of several motifs characteristic of a va
C:Genetics:
A:Gene: 54
A:Start codon: GTG
C:Superfamily: unassigned probable phosphoesterases; phosphoesterase core homology
```

```
Query Match 55.0%; Score 33; DB 2; Length 255;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLAKLLKG 9
    ||| :| :| :| :
Db 207 MKLAQYLKG 215

RESULT 12
S47091
cyclase-associated protein - Chlorohydra viridissima
C:Species: Chlorohydra viridissima
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 04-Sep-1998
C:Accession: S47091
R:Fenger, U.; Hofmann, M.; Galliot, B.; Schaller, H.C.
submitted to the EMBL Data Library, June 1994
A:Description: The role of the CAMP pathway in mediating the effect of head activator
A:Reference number: S47091
A:Accession: S47091
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-481 <FEN>
A:Cross-references: EMBL:X79567; NID:G498780; PID:G498781
C:Superfamily: adenyl cyclase-associated protein MCH1

Query Match 55.0%; Score 33; DB 2; Length 481;
Best Local Similarity 54.5%; Pred. No. 59;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 5 KLLKGEXXXIK 15
    ||| :| :| :| :
Db 182 QLLKGFAAYVK 192

RESULT 13
I55465
[pyruvate dehydrogenase (lipoamide)] kinase (EC 2.7.1.99) 1 - human
C:Species: Homo sapiens (man)
C:Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 04-Sep-1998
C:Accession: I55465
R:Gudi, R.; Bowker, M.M.; Kedishvili, N.Y.; Zhao, Y.; Popov, K.M.
J. Biol. Chem. 270, 28989-28994, 1995
A:Title: Diversity of the pyruvate dehydrogenase kinase gene family in humans.
A:Reference number: I55465
A:Accession: I55465
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-436 <RES>
A:Cross-references: GB:I42450; NID:G1088280; PID:G1088281
C:Genetics:
A:Gene: GDB:PDK1
A:Cross-references: GDB:3789442
C:Keywords: phosphotransferase

Query Match 55.0%; Score 33; DB 2; Length 436;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLAKLLKG 9
    ||| :| :| :| :
Db 1 MRLARLLRG 9

RESULT 14
A49686
[pyruvate dehydrogenase (lipoamide)] kinase (EC 2.7.1.99) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-May-1995 #sequence_revision 26-May-1995 #text_change 10-Sep-1997
C:Accession: A49686
R:Popov, K.M.; Kedishvili, N.Y.; Zhao, Y.; Shimomura, Y.; Crabb, D.W.; Harris, R.A.
```

J. Biol. Chem. 268, 26602-26606, 1993
A:Title: Primary structure of pyruvate dehydrogenase kinase establishes a new family of
A:Reference number: A49686; MUID:94075352
A:Accession: A49686
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-434 <POP>
A:Cross-references: GB:L22294; NID:g438873; PID:g438874
C:Keywords: mitochondrial matrix; mitochondrion; oxidoreductase; phosphotransferase
F:1-26/Domain: transit peptide (mitochondrion) #status predicted <TRP>

Query Match 55.08; Score 33; DB 2; Length 434;
Best Local Similarity 56.7%; Pred. No. 53;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLAKLLKG 9
|:|:|:|:|
Db 1 MRLARLLRG 9

RESULT 15
B71162
hypothetical protein PH0495 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence-revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: B71162
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137
A:Accession: B71162
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-440 <KAW>
A:Cross-references: GB:AP000002; NID:g3236129; PID:d1030526; PID:g3256900
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0495

Query Match 55.08; Score 33; DB 2; Length 440;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLAKLLKG 9
|:|:|:|:|
Db 180 KLAKMKG 187

Search completed: November 13, 1999, 12:08:04
Job time: 2076 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:33:39 ; Search time 51.07 Seconds
(without alignments)
8.303 Million cell updates/sec

Title: US-08-913-430-10

Perfect score: 60

Sequence: 1 MKLAKLLKGFXXIK 15

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	36	60.0	205	1	TIM1_MOUSE	P12032 mus musculus
2	36	60.0	217	1	TIM1_RAT	P30120 rattus norv
3	35	58.3	382	1	C79B_ARATH	O81346 arabidopsis
4	35	58.3	1022	1	DPO5_YEAST	P39985 saccharomyc
5	34	56.7	542	1	C79B_SINAL	O81345 sinapis alb
6	34	56.7	616	1	SPAN_STRPU	P98088 strongyloce
7	33	55.0	481	1	CAP_CHLVR	P40122 chlorohydra
8	33	55.0	89	1	MYSB_SPINA	P00118 spirulina m
9	33	55.0	1966	1	MYSB_CAEEL	P02566 caenorhabdi
10	33	55.0	436	1	PDK1_HUMAN	Q15118 homo sapien
11	33	55.0	434	1	PDK1_RAT	O63065 rattus norv
12	33	55.0	90	1	PTSO_ECOLI	P33996 escherichia
13	33	55.0	255	1	VG54_BPMD2	O84245 mycobacteri
14	33	55.0	255	1	VG54_BPML5	Q05264 mycobacteri
15	32	53.3	322	1	ASPG_BACLI	P30363 bacillus li
16	32	53.3	1161	1	DP3A_AQAE	O67125 quifex aeo
17	32	53.3	605	1	LEPA_TREPA	O83523 treponema p
18	32	53.3	476	1	OSTA_YEAST	P41543 saccharomyc
19	32	53.3	169	1	PNAAR_RICPR	P41077 rickettsia
20	32	53.3	207	1	TIM1_BOVIN	P20414 bos taurus
21	32	53.3	207	1	TIM1_HORSE	O02722 equus cabal
22	32	53.3	207	1	TIM1_PIG	P35624 sus scrofa
23	32	53.3	207	1	TIM1_SHEEP	P50122 ovis aries
24	32	53.3	379	1	UL28_HCMVA	P16847 human cytom
25	32	53.3	604	1	VEL_HPV38	Q80909 human papil
26	32	53.3	837	1	YL34_YEAST	Q07844 saccharomyc
27	32	53.3	269	1	YQXC_BAGSU	P19672 bacillus su
28	31	51.7	572	1	AA64_HUMAN	P29536 homo sapien
29	31	51.7	512	1	ANX7_XENLA	Q92125 xenopus lae
30	31	51.7	898	1	ATMA_ECOLI	P39168 escherichia
31	31	51.7	902	1	ATMA_SALTY	P36640 salmonella
32	31	51.7	442	1	BAM_DROME	P22745 drosophila
33	31	51.7	1062	1	CC7_SCHPO	P41892 schizosacch
34	31	51.7	4196	1	DYHC_SCHPO	Q13290 schizosacch
35	31	51.7	458	1	GSHR_ANASP	P48638 anabena sp
36	31	51.7	483	1	GSHR_YEAST	P41921 saccharomyc
37	31	51.7	926	1	KINH_NEUCR	P48467 neuropsora
38	31	51.7	724	1	PEDD_PEDAC	P36497 pedicoccus
39	31	51.7	743	1	RELA_HAEIN	P44644 haemophilus
40	31	51.7	114	1	REK20_GUTH	O78495 guillardia
41	31	51.7	117	1	RL20_SYNY3	P48957 synecocyst
42	31	51.7	339	1	RWAR_CANGA	P21358 candida gla
43	31	51.7	752	1	RRPI_INBAD	P13871 influenza b

ALIGNMENTS

```
RESULT 1
TIM1_MOUSE
ID TIM1_MOUSE STANDARD; PRT; 205 AA.
AC P12032: P20064:
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE METALLOPROTEINASE INHIBITOR 1 PRECURSOR (TIMP-1) (ERYTHROID
DE POTENTIATING ACTIVITY) (EPA) (TISSUE INHIBITOR OF METALLOPROTEINASES)
DE (COLLAGENASE INHIBITOR 16C8 FIBROBLAST) (TPA-INDUCED PROTEIN)
DE (TPA-SI).
GN TIMP1 OR TIMP-1 OR TIMP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87218524.
RA GEWERT D.R., COULOMBE B., CASTELINO M., SKUP D., WILLIAMS B.R.G.;
RT "Characterization and expression of a murine gene homologous to human
RL EPA/TIMP: a virus-induced gene in the mouse.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87066763.
RA EDWARDS D.R., WATERHOUSE P., HOLMAN M.L., DENHARDT D.T.;
RT "A growth-responsive gene (16C8) in normal mouse fibroblasts
homologous to a human collagenase inhibitor with
erythroid-potentiating activity: evidence for inducible and
constitutive transcripts.";
RL NUCLEIC ACIDS RES. 14:8863-8878(1986).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-C3H;
RX MEDLINE; 88038821.
RA JOHNSON M.D., HOUSBY G.M., KIRSCHMEIER P.T., WEINSTEIN I.B.;
RT "Molecular cloning of gene sequences regulated by tumor promoters and
mitogens through protein kinase C";
RL MOL. CELL. BIOL. 7:2821-2829(1987).
CC -!- FUNCTION: COMPLEXES WITH METALLOPROTEINASES (SUCH AS COLLAGENASES)
AND IRREVERSIBLY INACTIVATE THEM. ALSO MEDIATES ERYTHROPOIESIS IN
VITRO; BUT, UNLIKE IL-3, IT IS SPECIES-SPECIFIC, STIMULATING THE
GROWTH AND DIFFERENTIATION OF ONLY HUMAN AND MURINE ERYTHROID
PROGENITORS
CC -!- TISSUE SPECIFICITY: FOUND IN FETAL AND ADULT TISSUES. HIGHEST
LEVELS ARE FOUND IN BONE. ALSO FOUND IN LUNG, OVARY AND UTERUS.
CC -!- INDUCTION: BY VIRUS.
CC -!- INDUCTION: REGULATED BY TUMOR PROMOTERS AND MITOGENS THROUGH
PROTEIN KINASE C.
CC -!- PTM: THE ACTIVITY OF TIMP-1 IS DEPENDENT ON THE PRESENCE OF
DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO THE TIMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; M28312; G193042;
CC EMBL; M28308; G193042; JOINED.
CC EMBL; M28309; G193042; JOINED.
CC EMBL; M28310; G193042; JOINED.
CC
```



```
CC -!- FUNCTION: CONVERTS AMINO ACIDS TO ALDOXIMES IN GLUCOSINOLATE
CC BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF069495; G3283435; -.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC KW OXIDOREDUCTASE; MONOOXYGENASE; TRANSMEMBRANE; HEME.
CC FT BINDING 318 318 HEME (BY SIMILARITY).
CC FT BINDING 382 AA; 43521 MW; 4A96FA49 CRC32;
CC SEQUENCE 382 AA; 43521 MW; 4A96FA49 CRC32;
CC -----
CC Query Match 58.3%; Score 35; DB 1; Length 382;
CC Best Local Similarity 70.0%; Pred. No. 11;
CC Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
CC -----
Cc 1 MKLAKLKGKF 10
Cc 1 ||:||||
Cc 330 MMLARLLQGF 339
Cc -----
Cc RESULT 4
Cc ID DPOS_YEAST STANDARD; PRT; 1022 AA.
Cc AC P39985;
Cc DT 01-FEB-1995 (REL. 31, CREATED)
Cc DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
Cc DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
Cc DE DNA POLYMERASE V (EC 2.7.7.7) (POL V).
Cc GN POL5 OR YEL055C.
Cc OS SACCHAROMYCES CREVISIAE (BAKER'S YEAST).
Cc OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
Cc OC SACCHAROMYCETACEAE; SACCHAROMYCES.
Cc RN [1]
Cc RP SEQUENCE FROM N.A.
Cc RC STRAIN-S288C / AB972;
Cc RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
Cc RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
Cc RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,
Cc RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
Cc RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,
Cc RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOEGREN T., SMITH V.,
Cc RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;
Cc RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
Cc RN [2]
Cc RP IDENTIFICATION.
Cc RA MEDLINE; 95397432.
Cc RA SUGINO A.;
Cc RT "Yeast DNA polymerases and their role at the replication fork.";
Cc RL TRENDS BIOCHEM. SCI. 20:319-323(1995).
Cc CC -!- FUNCTION: PRESUMED TO PLAY A ROLE OTHER THAN CHROMOSOMAL DNA
Cc REPLICATION.
Cc CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
Cc N PYROPHOSPHATE + DNA(N).
Cc CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
Cc -----
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb-sib.ch/announcement/
Cc or send an email to license@isb-sib.ch).
Cc -----
Cc EMBL; U18795; G603263; -.
Cc DR
```

```
DR SGD; L0004119; POL5.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW TRANSFERASE; DNA-DIRECTED DNA POLYMERASE.
SQ SEQUENCE 1022 AA; 115893 MW; A5FD652C CRC32;
CC -----
Cc Query Match 58.3%; Score 35; DB 1; Length 1022;
Cc Best Local Similarity 57.1%; Pred. No. 29;
Cc Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Cc -----
Cc 2 KLAKLLKGFFXXIK 15
Cc 1:::|||||
Cc Db 906 KISKLLKGKIFKIK 919
Cc -----
Cc RESULT 5
Cc ID C79B_SINAL STANDARD; PRT; 542 AA.
Cc AC 081345;
Cc DT 15-DEC-1998 (REL. 37, CREATED)
Cc DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
Cc DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
Cc DE CYTOCHROME P450 79B1 (EC 1.14.-.-).
Cc GN CYP79B1.
Cc OS SINAPIS ALBA (WHITE MUSTARD) (BRASSICA HIRT).
Cc OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
Cc OC EUPHYLLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
Cc OC CAPPARALES; BRASSICACEAE; SINAPIS.
Cc RN [1]
Cc RP SEQUENCE FROM N.A.
Cc RC STRAIN-CV. CARLA;
Cc RA BAK S., NIELSEN H.L., HALKIER B.A.;
Cc RT "The presence of cyp79 homologues in glucosinolate-producing plants
Cc shows evolutionary conservation of the enzymes in the conversion of
Cc amino acids to aldoxime in the biosynthesis of cyanogenic glucosides
Cc and glucosinolates.";
Cc RL PLANT MOL. BIOL. 0:0-0(1998).
Cc CC -!- FUNCTION: CONVERTS TYROSINE TO PARA-HYDROPHENYLACETALDOXIME IN
Cc PARA-HYDROXYBENZYLGLUCOSINOLATE BIOSYNTHESIS.
Cc CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
Cc -----
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb-sib.ch/announcement/
Cc or send an email to license@isb-sib.ch).
Cc -----
Cc EMBL; AF069494; G3283433; -.
Cc DR PROSITE; PS00086; CYTOCHROME_P450; 1.
Cc KW OXIDOREDUCTASE; MONOOXYGENASE; TRANSMEMBRANE; HEME.
Cc FT TRANSMEM 21 41 POTENTIAL.
Cc FT BINDING 478 478 HEME (BY SIMILARITY).
Cc SEQUENCE 542 AA; 61291 MW; 7996FC8D CRC32;
Cc -----
Cc Query Match 56.7%; Score 34; DB 1; Length 542;
Cc Best Local Similarity 70.0%; Pred. No. 24;
Cc Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Cc -----
Cc 1 MKLAKLKGKF 10
Cc 1:::|||||
Cc Db 490 MMLARLLQGF 499
Cc -----
Cc RESULT 6
Cc ID SPAN_STRPU STANDARD; PRT; 616 AA.
Cc AC P98058;
Cc DT 01-FEB-1996 (REL. 33, CREATED)
Cc DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
Cc DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
```



```

DR HSP; P55534; IC6S.
KW ELECTRON TRANSPORT; PHOTOSYNTHESIS; HEME.
FT BINDING 14 HEME (COVALENT).
FT BINDING 17 HEME (COVALENT).
FT METAL 18 IRON (HEME AXIAL LIGAND).
FT METAL 62 IRON (HEME AXIAL LIGAND).
SQ SEQUENCE 89 AA; 9236 MW; 84210796 CRC32;

Query Match 55.0%; Score 33; DB 1; Length 89;
Best Local Similarity 87.5%; Pred. No. 6.5;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 LAKLKGK 10
    ||| ||||
Db 36 LAKYLKGF 43

RESULT 9
MSB_CAEEL STANDARD; PRT; 1966 AA.
AC P02566;
DT 21-JUL-1986 (REL. 01, LAST CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN B (MHC B).
GN UNC-54 OR MYO-4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
[1]
SEQUENCE FROM N.A.
RN MEDLINE; 83273600.
RA KARN J., BRENNER S., BARNETT L.;
RT "Protein structural domains in the Caenorhabditis elegans unc-54
  myosin heavy chain gene are not separated by introns.";
RL PROC. NATL. ACAD. SCI. U.S.A. 80:4253-4257(1983).
[2]
SEQUENCE OF 850-1966 FROM N.A.
RN MEDLINE; 83232892.
RA WILLS N., GESTELAND R.F., KARN J., BARNETT L., BOLTON S.,
  WATERSTON R.H.;
RT "The genes sup-7 x and sup-5 III of C. elegans suppress amber
  nonsense mutations via altered transfer RNA.";
RL CELL 33:575-583(1983).
[3]
SEQUENCE OF 1876-1966 FROM N.A.
RN MEDLINE; 83232892.
RA WILLS N., GESTELAND R.F., KARN J., BARNETT L., BOLTON S.,
  WATERSTON R.H.;
RT "The genes sup-7 x and sup-5 III of C. elegans suppress amber
  nonsense mutations via altered transfer RNA.";
RL CELL 33:575-583(1983).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
  HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
  AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
  CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
  CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
  ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMM)
  AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE SPLIT FURTHER INTO
  2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C. ELEGANS.
  THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
CC -1- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
  SIMILARITY WITH THE GLOBULAR HEAD AA SEQUENCES OF MUSCLE & NONMUSCLE
  HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
  THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
  THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J01050; G156400; -.
DR EMBL; V01494; G6784; -.
DR FIR; A02992; MKW.
DR FRAM; PF00063; myosin_head; 1.
DR HSP; P08799; 1MND.
KW MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;
KW ATP-BINDING; METHYLATION; ALKYLATION; HEPTAD REPEAT PATTERN;
KW MULTIGENE FAMILY.
FT DOMAIN 1 850 GLOBULAR HEAD (S1).
FT DOMAIN 851 1966 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 851 1164 ALPHA-HELICAL TAIPIECE (S2).
FT DOMAIN 1165 1966 LIGHT MEROMYOSIN (LMM).
FT DOMAIN 851 1966 COILED COIL (POTENTIAL).
FT NP_BIND 177 184 ATP (BY SIMILARITY).
FT DOMAIN 665 687 ACTIN-BINDING.
FT DOMAIN 769 783 ACTIN-BINDING.
FT MOD_RES 128 128 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 705 705 ALKYLATION (SH-1).
FT MOD_RES 715 715 ALKYLATION (SH-2).
FT DOMAIN 1165 1176 HINGE.
FT CONFLICT 1337 1337 E -> R (IN REF. 2).
FT CONFLICT 1880 1880 I -> L (IN REF. 2).
SQ SEQUENCE 1966 AA; 225125 MW; C59A26E4 CRC32;

Query Match 55.0%; Score 33; DB 1; Length 1966;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 KLAKLLGFXKXIK 15
    ||| :| :| :|
Db 792 KLATILTGFSQIR 805

RESULT 10
PDK1_HUMAN STANDARD; PRT; 436 AA.
AC Q15118;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE [PYRUVATE DEHYDROGENASE(LIPOAMIDE)] KINASE ISOZYME 1 PRECURSOR
DE (EC 2.7.1.99) (PYRUVATE DEHYDROGENASE KINASE ISOFORM 1).
GN PDK1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE; 96081973.
RA GUDI R., BOWKER-KINLEY M.M., KEDISHVILI N.Y., ZHAO Y., POPOV K.M.;
RT "Diversity of the pyruvate dehydrogenase kinase gene family in
  humans.";
RL J. BIOL. CHEM. 270:28989-28994(1995).
CC -1- FUNCTION: INHIBITS THE MITOCHONDRIAL PYRUVATE DEHYDROGENASE
  COMPLEX BY PHOSPHORYLATION OF THE E1 ALPHA SUBUNIT, THUS
  CONTRIBUTING TO THE REGULATION OF GLUCOSE METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + [PYRUVATE DEHYDROGENASE (LIPOAMIDE)] =
  ADP + [PYRUVATE DEHYDROGENASE (LIPOAMIDE)] PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE HEART.
CC -1- SIMILARITY: BELONGS TO THE PDK/BCKDK FAMILY OF PROTEIN KINASES.
CC -----

```


This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: L42450; G1088281; -
 DR MIN: 602524; -
 KW KINASE; TRANSFERASE; TRANSIT PEPTIDE; MITOCHONDRION; MULTIGENE FAMILY.
 FT CHAIN 1 28 MITOCHONDRION (POTENTIAL).
 FT CHAIN 29 436 PYRUVATE DEHYDROGENASE KINASE ISOFORM 1.
 SQ SEQUENCE 436 AA; 49244 MW; 52DC3006 CRC32;

Query Match 55.08; Score 33; DB 1; Length 436;
 Best Local Similarity 66.7%; Pred. No. 31;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLAKLKG 9
 Db 1 MRLARLLRG 9

RESULT 11
 PDK1_RAT
 ID PDK1_RAT STANDARD; PRT; 434 AA.
 AC Q63055;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DE (PYRUVATE DEHYDROGENASE(LIPOAMIDE)) KINASE ISOZYME 1 PRECURSOR
 DE (EC 2.7.1.99) (PYRUVATE DEHYDROGENASE KINASE ISOFORM 1) (PDK P48).
 GN PDK1 OR PDH.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 RC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-HEART;
 RX MEDLINE; 94075332.
 RA POPOV K.M., KEDISHVILI N.Y., ZHAO Y., SHIMOMURA Y., CRABB D.W.,
 RA HARRIS R.A.;
 RT "Primary structure of pyruvate dehydrogenase kinase establishes a new family of eukaryotic protein kinases.";
 RL J. BIOL. CHEM. 268:26602-26606(1993).
 CC -!- FUNCTION: INHIBITS THE MITOCHONDRIAL PYRUVATE DEHYDROGENASE COMPLEX BY PHOSPHORYLATION OF THE E1 ALPHA SUBUNIT, THUS CONTRIBUTING TO THE REGULATION OF GLUCOSE METABOLISM.
 CC -!- CATALYTIC ACTIVITY: ATP + (PYRUVATE DEHYDROGENASE (LIPOAMIDE)) - ADP + (PYRUVATE DEHYDROGENASE (LIPOAMIDE)) PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE HEART.
 CC -!- SIMILARITY: BELONGS TO THE PDK/CKDK FAMILY OF PROTEIN KINASES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: L22294; G43874; -
 KW KINASE; TRANSFERASE; TRANSIT PEPTIDE; MITOCHONDRION; MULTIGENE FAMILY.
 FT CHAIN 1 26 MITOCHONDRION (POTENTIAL).
 FT CHAIN 27 434 PYRUVATE DEHYDROGENASE KINASE ISOFORM 1.
 SQ SEQUENCE 434 AA; 49081 MW; A5C1C43B CRC32;

Query Match 55.08; Score 33; DB 1; Length 434;
 Best Local Similarity 66.7%; Pred. No. 31;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MKLAKLKG 9
 Db 1 MRLARLLRG 9

RESULT 12
 PTSO_ECOLI
 ID PTSO_ECOLI STANDARD; PRT; 90 AA.
 AC P33996;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE PHOSPHOCARRIER PROTEIN NPR (NITROGEN RELATED HPR).
 GN PTSO OR NPR OR RPOR.
 OS ESCHERICHIA COLI.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ESCHERICHIA.
 [1]
 RN RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN-K12 / W3110;
 RX MEDLINE; 95181483.
 RA POWELL B.S., COURT D.L., INADA T., NAKAMURA Y., MICHOTY V.,
 RA CUI X., REIZER A., SAIER M.H. JR., REIZER J.;
 RT "Novel proteins of the phosphotransferase system encoded within the rpoN operon of Escherichia coli. Enzyme IANtr affects growth on organic nitrogen and the conditional lethality of an erats mutant.";
 RL J. BIOL. CHEM. 270:4822-4833(1995).
 [2]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE; 94297724.
 RA JONES D.H.A., FRANKLIN C.F.H., THOMAS C.M.;
 RT "Molecular analysis of the operon which encodes the RNA polymerase sigma factor sigma 54 of Escherichia coli.";
 RL MICROBIOLOGY 140:1035-1043(1994).
 [3]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA PLUNKETT G. III;
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- FUNCTION: SEEMS TO HAVE A ROLE IN LINKING CARBON AND NITROGEN ASSIMILATION. PROBABLY ACT IN A REGULATORY CAPACITY AND COULD CONTROL THE STATE OF PHOSPHORYLATION OF IIA-NTR (PTSN).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
 CC -!- PTM: PROBABLY PHOSPHORYLATED BY A YET UNCHARACTERIZED ENZYME I.
 CC -!- SIMILARITY: TO ALL OTHER HPR OR HPR DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: U12684; G551342; -
 DR EMBL: U12684; G551342; -
 DR EMBL: U18997; G606145; -
 DR EMBL: AEO00400; G1789599; -
 DR PIR: S38619; S38619.
 DR ECOGENE: EG12147; PTSO.
 DR PROSITE; PS00369; PTS_HPR_HIS; 1.
 DR PROSITE; PS00589; PTS_HPR_SER; 1.
 DR PFAM; PF00381; PTS_HPR; 1.
 DR HSP; P08877; 2HPR.
 KW PHOSPHOTRANSFERASE SYSTEM; PHOSPHORYLATION.
 FT MOD_RES 16 16 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 48 48 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 90 AA; 9810 MW; C7558A8A CRC32;

```
Query Match      55.0%; Score 33; DB 1; Length 90;
Best Local Similarity 42.9%; Pred. No. 6.5;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MKLAKLLKGFXXIXI 14
DB 21 MKLFELMOGFDREV 34

RESULT 13
VG54_BPMD2      STANDARD;      PRT; 255 AA.
AC 064245;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENE 54 PROTEIN (GP54).
GN 54
OS MYCOBACTERIOPHAGE D29.
OC VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98300335.
RA FORD M.E., SARKIS G.J., BELANGER A.E., HENDRIX R.W., HATFULL G.F.;
RT "Genome structure of mycobacteriophage D29: implications for phage
  evolution."
RL J. MOL. BIOL. 279:143-164(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF022214; G3172302;
SQ SEQUENCE 255 AA; 28546 MW; 918977A2 CRC32;

Query Match      55.0%; Score 33; DB 1; Length 255;
Best Local Similarity 77.8%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLAKLLKG 9
DB 207 MKLAQYLKG 215

RESULT 14
VG54_BPML5      STANDARD;      PRT; 255 AA.
AC 005264;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE GENE 54 PROTEIN (GP54).
GN 54
OS MYCOBACTERIOPHAGE L5.
OC VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93211282.
RA HATFULL G.F., SARKIS G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
  a phage system for mycobacterial genetics."
RL MOL. MICROBIOL. 7:395-405(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF022214; G3172302;
SQ SEQUENCE 255 AA; 28546 MW; 918977A2 CRC32;

Query Match      55.0%; Score 33; DB 1; Length 255;
Best Local Similarity 77.8%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKLAKLLKG 9
DB 207 MKLAQYLKG 215

RESULT 15
ASPG_BACLI      STANDARD;      PRT; 322 AA.
ID ASPG_BACLI
AC P30363;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE L-ASPARAGINASE (EC 3.5.1.1) (L-ASPARAGINE AMIDOHYDROLASE).
GN ANSA.
OS BACILLUS LICHENIFORMIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92009126.
RA VAN DIJL J.M., DE JONG A., SMITH H., BRON S., VENEMA G.;
RT "Lack of specific hybridization between the lep genes of Salmonella
  typhimurium and Bacillus licheniformis."
RL FEMS MICROBIOL. LETT. 65:345-351(1991).
CC -1- CATALYTIC ACTIVITY: L-ASPARAGINE + H(2)O -> L-ASPARTATE + NH(3).
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11497; G49272;
SQ SEQUENCE 322 AA; 35442 MW; 11F86F29 CRC32;

Query Match      53.3%; Score 32; DB 1; Length 322;
Best Local Similarity 53.3%; Pred. No. 36;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKLAKLLKGFXXIXI 15
DB 303 MKLAVLLASYKEGK 317

Search completed: November 13, 1999, 10:33:41
Job time: 5191 sec
```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:34 ; Search time 139.86 Seconds
(without alignments)
6.601 Million cell updates/sec

Title: US-08-913-430-10

Perfect score: 60

Sequence: 1 MKLAKLLKGFXXXIK 15

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

- SPTREMBL10:**
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	61.7	2166	2	O51465 borrelia bu
2	36	60.0	904	2	O50187 mycoplasma
3	35	60.0	176	11	P70533 ratius norv
4	35	58.3	109	1	O58934 methanococc
5	35	58.3	173	1	O59104 pyrococcus
6	34	56.7	913	3	Q12151 saccharomyc
7	34	56.7	481	10	Q92WJ3 arabidopsis
8	33	55.0	440	1	O58231 pyrococcus
9	33	55.0	435	2	O52562 amycolatops
10	33	55.0	1963	5	O02244 caenorhabdi
11	32	53.3	361	1	O57936 pyrococcus
12	32	53.3	133	2	O25418 helicobacte
13	32	53.3	266	2	O31424 bacillus su
14	32	53.3	146	2	O87096 bacillus su
15	32	53.3	133	2	Q921C6 helicobacte
16	32	53.3	312	3	Q12125 saccharomyc
17	32	53.3	312	3	Q92264 pichia past
18	32	53.3	618	3	O60134 schizosacch
19	32	53.3	579	3	Q06820 saccharomyc
20	32	53.3	26926	4	O10466 homo sapien
21	32	53.3	391	4	O95397 homo sapien
22	32	53.3	907	5	Q26675 theileria a
23	32	53.3	296	5	O02356 caenorhabdi
24	32	53.3	2496	5	O96175 plasmodium
25	32	53.3	6875	6	Q28733 oryctolagus
26	32	53.3	175	10	Q92S24 triticum ae
27	32	53.3	179	10	Q92S23 triticum ae
28	32	53.3	605	12	O58954 human papil
29	31	51.7	328	1	O58765 methanococc

ALIGNMENTS

RESULT 1

O51465
ID O51465 PRELIMINARY; PRT; 2166 AA.
AC O51465;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HYPOTHEICAL 254.2 KD PROTEIN.
GN BB0512.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE; 98065943.
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi".
RL Nature 390:580-586(1997).
DR EMBL; AE001153; AAC66876.1; -.
DR TIGR; BB0512; -.
KW Hypothetical protein.
SQ SEQUENCE 2166 AA; 254243 MW; F2899A0A CRC32;

Query Match 61.7%; Score 37; DB 2; Length 2166;
Best Local Similarity 61.5%; Pred. No. 1.2e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 KLAKLLKGFXXXI 14

Db 131 KLSKTLKGFTQI 143

RESULT 2

O50187
ID O50187 PRELIMINARY; PRT; 904 AA.
AC O50187;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
DE P102.
GN Mycoplasma hyopneumoniae.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Mycoplasmatocaceae; Mycoplasma.

RN {1}
RP SEQUENCE FROM N.A.
RC STRAIN-232A;
RA HSU T., MINION F.C.; to the EMBL/GenBank/DBJ databases.
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF012905; AAC32527.1; -.
SQ SEQUENCE 904 AA; 102343 MW; 870D5B48 CRC32;

Query Match 60.08; Score 36; DB 2; Length 904;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLAKLLK 8

Db 1 MKLAKLLK 8

RESULT 3

P70533 PRELIMINARY; PRT; 176 AA.

AC P70533;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE METALLOPROTEINASE INHIBITOR (FRAGMENT).
GN TIMP-1.
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]

RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;

RX MEDLINE; 96317624.

RA IREDALE J.P., CLARK N., MURPHY G.;

RA "Tissue inhibitor of metalloproteinase-1 messenger RNA expression is enhanced relative to interstitial collagenase messenger RNA in experimental liver injury and fibrosis.";
RL Hepatology 24:176-184(1996).

DR EMBL; L29512; AAB08483.1; -.

DR PFAM; PF00965; TIMP; 1.

FT NON_TER 1

SQ SEQUENCE 176 AA; 19659 MW; 7CB8A5D1 CRC32;

Query Match 60.08; Score 36; DB 11; Length 176;
Best Local Similarity 60.0%; Pred. No. 15;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKLAKLLKGF 10

Db 34 IKWTKMLKGF 43

RESULT 4

Q58934 PRELIMINARY; PRT; 109 AA.

AC Q58934;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TReMBLrel. 06, Last annotation update)
DE HYPOTHETICAL PROTEIN M31539.
GN M31539.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;

OC Methanococcus.

[1]

RN. SEQUENCE FROM N.A.

RC STRAIN-DSM 2661.

RX MEDLINE; 96337999.

RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,

RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
RA SCOTT J.L., GEOCHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOSE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO A.FULGIDUS AF1885.
DR EMBL; U67595; AAB99568.1; -.
KW Hypothetical protein.
SQ SEQUENCE 109 AA; 12451 MW; 386B81CB CRC32;

Query Match 58.3%; Score 35; DB 1; Length 109;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 KLAKLLKGXXXIK 15

Db 71 KLAKILGGSGSEVR 84

RESULT 5

O59104 PRELIMINARY; PRT; 173 AA.

ID O59104;

DT 01-AUG-1998 (TReMBLrel. 07, Created)

DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)

DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)

DE 173AA LONG HYPOTHETICAL NADH-UBIQUINONE OXIDOREDUCTASE SUBUNIT.

GN PH1434.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

[1]

RP SEQUENCE FROM N.A.

RX STRAIN=OT3;

RX MEDLINE; 98344137.

RA KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HATAKAWA Y., HINO Y.,

RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,

RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,

RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,

RA AKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,

RA KIKUCHI H.;

RT "Complete sequence and gene organization of the genome of a

hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

RL DNA Res. 5:55-76(1998).

DR EMBL; AP000006; BAA30541.1; -.

DR PFAM; PF01058; oxidored_q6; 1.

KW Ubiquinone.

SQ SEQUENCE 173 AA; 18895 MW; AB231FE8 CRC32;

Query Match 58.3%; Score 35; DB 1; Length 173;
Best Local Similarity 87.5%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KLAKLLKG 9

Db 157 KLAKMLKG 164

RESULT 6

Q12151 PRELIMINARY; PRT; 913 AA.

ID Q12151;

AC Q12151;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)

DE HYPOTHETICAL 100.3 KD PROTEIN YD8142.

GN YD8142.14 OR YD8142B.05.

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB972;
 RA OLIVER K., SHORE L., HARRIS D.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB972;
 RA BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AB972;
 RA BARRELL B., RAJANDREAM M.A., WALSH S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
 CLUSTER DOMAIN.
 DR EMBL; 268194; CAA92356.1; -;
 DR EMBL; 268195; CAA92364.1; -;
 DR PFAM; PF00172; Zn_clus.1;
 DR PROSITE; PS00463; ZN2_Cy6_FUNGAL_1; 1.
 KW Hypothetical protein; Transcription regulation; DNA-binding;
 KW Nuclear protein; Zinc; Metal-binding.

SQ SEQUENCE 913 AA; 100339 MW; 620E7F72 CRC32;
 Query Match 56.7%; Score 34; DB 3; Length 913;
 Best Local Similarity 54.5%; Pred. NO. 1.8e+02;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 KLLGCFXXXIK 15
 |||:|:|:|
 DB 853 KLLRGFAFEVK 863

RESULT 7
 Q92WJ3 PRELIMINARY; PRT; 481 AA.
 AC Q92WJ3;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE UDP-GLUCOSE GLUCOSYLTRANSFERASE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLOMBIA;
 RA YAMAGISHI E., GOND Z., YAMAZAKI M., SAITO K.;
 RT "Molecular cloning of a cDNA encoding a novel VDP-Glucose
 glucosyltransferase homologue from Arabidopsis thaliana (Accession
 No. AB016819). (PCR98-187)."
 RL Plant Physiol. 118:1102-1102(1998).
 DR EMBL; AB016819; BAA34687.1; -;
 DR PROSITE; PS00375; UDPGT; 1.
 KW Transferase.
 SQ SEQUENCE 481 AA; 54364 MW; C287B725 CRC32;

Query Match 56.7%; Score 34; DB 10; Length 481;
 Best Local Similarity 75.0%; Pred. NO. 98;
 Matches 9; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 1 MKLAKLL--KGF 10
 ||:|||||
 DB 26 MKVAKLLYAKGF 37

Query Match 56.7%; Score 34; DB 10; Length 481;
 Best Local Similarity 75.0%; Pred. NO. 98;
 Matches 9; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 1 MKLAKLL--KGF 10
 ||:|||||
 DB 26 MKVAKLLYAKGF 37

RESULT 8

O58231 PRELIMINARY; PRT; 440 AA.
 AC O58231;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
 DE 440AA LONG HYPOTHETICAL PROTEIN.
 GN PH0495.
 OS Pyrococcus horikoshii.
 CC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE; 98344137.
 RA KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOIYAMA A., NAGAI Y.,
 RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
 RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
 RA KIKUCHI H.;
 RT "Complete sequence and gene organization of the genome of a
 hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
 RT DNA Res. 5:55-76(1998).
 DR EMBL; AP000002; BAA29583.1; -;
 SQ SEQUENCE 440 AA; 47374 MW; 776F69BE CRC32;

Query Match 55.0%; Score 33; DB 1; Length 440;
 Best Local Similarity 75.0%; Pred. NO. 1.4e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLAKLLKG 9
 |||:|:|
 DB 180 KLAKMIRG 187

RESULT 9

O52562 PRELIMINARY; PRT; 435 AA.
 AC O52562;
 DT 01-JUN-1998 (TrEMBLrel. 05, Created)
 DT 01-JUN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE Dntp-HEXOSE DEHYDRATASE.
 OS Amycolatopsis mediterranei.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae;
 OC Amycolatopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S699;
 RA KIM C.G., YU T.W., FRYHLE C., HANDA S., FLOSS H.G.;
 RA J. Biol. Chem. 0:0-0(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S699;
 RA AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., YU T.W.,
 RA TAYLOR M., HOFFMANN D., KIM C.G., ZHANG X., HUTCHINSON C.R.,
 RA FLOSS H.G.;
 RL Chem. Biol. 5:0-0(0002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S699;
 RA AUGUST P.R., TANG L., YOON Y.J., NING S., MUELLER R., HUTCHINSON C.R.,
 RA FLOSS H.G.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF040570; AAC01730.1; -;
 DR PFAM; PF01041; Degt_DnrJ_EryCI; 1.
 SQ SEQUENCE 435 AA; 47465 MW; 9452355F CRC32;

```
Query Match          55.0%; Score 33; DB 2; Length 435;
Best Local Similarity 70.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLAKLLKGF 10
Db 235 IKLAKLSKF 244

RESULT 10
O02244
ID O02244 PRELIMINARY; PRT; 1963 AA.
AC O02244;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE UNC-54 PROTEIN.
GN UNC-54.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA KERSHAW J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WORLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Nature 368:32-38(1994).
DR EMBL; 283107; CAB05505.1; -.
DR EMBL; 281499; CAB05505.1; JOINED.
DR EMBL; 281499; CAB04089.1; -.
DR EMBL; 283107; CAB04089.1; JOINED.
DR PFAM; PF00063; myosin_head; 1.
SQ SEQUENCE 1963 AA; 224754 MW; DACC0BF9 CRC32;

Query Match          55.0%; Score 33; DB 5; Length 1963;
Best Local Similarity 50.0%; Pred. No. 6.1e+02;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 KLAKLLKGFXXXIK 15
Db 789 KLATILTFQSOIR 802

RESULT 11
O57936
ID O57936 PRELIMINARY; PRT; 361 AA.
AC O57936;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE 361AA LONG HYPOTHETICAL PROTEIN.
GN PH0197.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
```

DE YBCQ PROTEIN.
 GN Bacillus subtilis.
 OS Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE; 98044033.
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
 RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
 RA GHIM S.Y., GLASER P., GOFEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GUISEPPI G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A.,
 RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
 RA JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
 RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 RA NOONE D., O'REILLY M., OGAWA K., OGIMURA A., OUDEGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTELELLA D., PORWOLLIK S., PRESCOTT A.M.,
 RA PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCORFONE F.,
 RA SERIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDÓ B.,
 RA SOROKIN A., TACCONEI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
 RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
 RA WINTERS P., WIRAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256(1997).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z99105; CAB11986.1; -.
 DR PFAM; PF01444; MoaA_NiFB_PqqE; 1.
 SQ SEQUENCE 266 AA; 29401 MW; 99830E41 CRC32;

Query Match 53.3%; Score 32; DB 2; Length 266;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 AKLLKGF 10
 III:III
 DB 19 AKLIKGF 25

RESULT 14
 ID O87096 PRELIMINARY; PRT; 146 AA.
 AC O87096;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
 DE YBCQ PROTEIN.
 GN YBCQ.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RC SEQUENCE FROM N.A.

RC STRAIN-168;
 RA HAGA K., LIU H., YASUMOTO K., TAKAHASHI H., YOSHIKAWA H.;
 RT "Sequence analysis of the 70kb region between 17 and 23 degree of the
 RT Bacillus subtilis chromosome.";
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB006424; BAA33088.1; -.
 DR PFAM; PF01444; MoaA_NiFB_PqqE; 1.
 SQ SEQUENCE 146 AA; 15983 MW; 397C19E8 CRC32;

Query Match 53.3%; Score 32; DB 2; Length 146;
 Best Local Similarity 85.7%; Pred. No. 73;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 AKLLKGF 10
 III:III
 DB 19 AKLIKGF 25

RESULT 15
 ID Q9ZLC6 PRELIMINARY; PRT; 133 AA.
 AC Q9ZLC6;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
 DE PUTATIVE.
 GN JHP0654.
 OS Helicobacter pylori J99.
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-J99;
 RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
 RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
 RA TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
 RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
 RA TRUST T.J.;
 RT "Genomic Sequence Comparison of Two Unrelated Isolates of the Human
 RT Gastric Pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 DR EMBL; AB001497; AAD06230.1; -.
 SQ SEQUENCE 133 AA; 15033 MW; 6C52497F CRC32;

Query Match 53.3%; Score 32; DB 2; Length 133;
 Best Local Similarity 50.0%; Pred. No. 67;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 KLAKLLKGFXXIK 15
 II:II:II
 DB 102 KLEKILKYDLAIK 115

Search completed: November 13, 1999, 12:55:35
 Job time: 3034 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:59:57 ; Search time 104.22 Seconds
(without alignments)
3.409 Million cell updates/sec

Title: US-08-913-430-11

Perfect score: 58

Sequence: 1 ADPXXYPQGXV 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	46	79.3	15	W01029	Mycoplasma 60-64 k
2	35	60.3	529	W80498	A protein designat
3	35	60.3	529	W85055	Human protein desi
4	33	56.9	329	R44997	Mob polypeptide en
5	31	53.4	38	R58479	TSAR binding domai
6	30	51.7	243	R58815	Human c-myc far up
7	30	51.7	590	R58816	Human c-myc far up
8	30	51.7	644	R58813	Human c-myc far up
9	30	51.7	643	R58814	Human c-myc far up
10	30	51.7	1291	W59912	Amino acid sequenc
11	30	51.7	189	W72395	Pathogen response
12	29	50.0	441	R31955	Sequence encoded b
13	29	50.0	161	R85475	PTLV-L Rex protein
14	29	50.0	516	R77522	Cytochrome-P450 CS
15	29	50.0	516	R77511	Cytochrome-P450-lp
16	29	50.0	441	R79035	Infectious bovine
17	29	50.0	799	W35293	Human disintegrin
18	29	50.0	749	W56134	Mus musculus K02 P
19	29	50.0	748	W56132	Homo sapiens trans
20	29	50.0	799	W70456	Human disintegrin
21	29	50.0	748	W70457	Mutant human disin
22	29	50.0	502	W80398	A secreted protein
23	29	50.0	305	W73519	KSHV assembly prot
24	29	50.0	300	W73518	KSHV assembly prot
25	28	48.3	455	R12362	Octopus rhodopsin
26	28	48.3	2396	R29939	Deduced from Lelys
27	28	48.3	1308	R54841	HER4. New recombin
28	28	48.3	541	R54842	HER4 N-terminal tr
29	28	48.3	269	R60558	Human basigin I. H
30	28	48.3	503	R72363	Human cytochrome P
31	28	48.3	1308	R91733	Receptor tyrosine
32	28	48.3	541	R91736	Receptor tyrosine
33	28	48.3	864	R93021	Human glucagon deg
34	28	48.3	503	R81464	Human derived cyto
35	28	48.3	503	R93170	Human cytochrome P
36	28	48.3	180	W13579	Batten disease cLN
37	28	48.3	705	W79144	Receptor protein t
38	28	48.3	101	W72387	Pathogen response
39	28	48.3	705	W81401	Receptor protein t
40	28	48.3	389	W73165	S. aureofaciens te
41	28	48.3	24	W95022	Sorbitol dehydroge
42	28	48.3	763	Y03219	Amino acid sequenc
43	28	48.3	46	Y11790	Human 5' EST secre

ALIGNMENTS

RESULT 1

W01029 W01029 standard; Peptide; 15 AA.
ID W01029;
DT 19-JAN-1997 (first entry)
DE Mycoplasma 60-64 kDa protective antigen N-terminal peptide.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumonia;
KW diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
FH Key
FT misc_difference 4 Location/Qualifiers
FT /label= Phe, Ile
FT /label= Arg, Glu
FT /label= Val, Ala
FT /label= Gln, Ala
FT /note= "undetermined amino acid"
FT misc_difference 12
FT misc_difference 13 /label= Met, Asn
PN W09628472-Al.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (UYME) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
DR WPI; 96-433763/43.
PT Putative protective antigens against Mycoplasma - used for the
PT detection, prevention or treatment of Mycoplasma infections, esp. M.
PT hyopneumoniae in swine
PS Claim 9; Page 27; 43pp; English.
CC Two 60-64 kDa putative protective antigens against Mycoplasma
CC were identified that respectively contain the N-terminal sequences
CC given in W01028 and W01029. The antigens were isolated from
CC Mycoplasma hyopneumoniae cells using antibody probes enriched with
CC mycoplasma-specific antibodies. Other protective antigens were
CC also identified (see also W01024-27 and W01030-37). Protective
CC antigens and antibodies can be used in vaccines for preventing or
CC treating mycoplasma infections, partic. M. hyopneumoniae
CC infections in swine. They can also be used for diagnosis.
SQ Sequence 15 AA;

Query Match 79.3%; Score 46; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADPXXYPQGXV 15

Db 1 ADPXXYPQGXV 15

RESULT 2

W80498 W80498 standard; Protein; 529 AA.
ID W80498;
DT 10-FEB-1999 (first entry)
DE A protein designated corneodesmosin.
KW Corneodesmosin; intercorneocyte cohesion; skin trophic disorder;
KW scar formation; epidermis thinning; skin fragility; hyperkeratosis;
KW xerosis; ichthyosis; psoriasis; hyperkeratotic tumour;
KW reactive keratosis; and leucokeratosis.
OS Homo sapiens.
PN FR2761362-Al.

PD 02-OCT-1998.
 PF 28-MAR-1997; 003899.
 PR 28-MAR-1997; FR-003899.
 PA (OREA) L'OREAL SA.
 PI Serre G, Simon M, Weber VM;
 DR WPI; 98-586093/50.
 DR N-PSDB; V63614.
 PT New polypeptide, corneodesmosin, and related nucleic acid - for
 PT cosmetic or therapeutic treatment of trophic disorders or thinning
 PT of the epidermis
 PS Claim 1; Page 29; 35pp; French.
 CC The present sequence represents a protein designated corneodesmosin.
 CC The protein is involved in intercorneocyte cohesion and is used, in
 CC cosmetic or pharmaceutical compositions, to treat trophic disorders
 CC of the skin and disorders associated with scar formation, and
 CC thinning of the epidermis, particularly the corneal layer, or skin
 CC fragility, to reinforce intercorneocyte cohesion and/or to induce
 CC thickening of the corneal layer. Hyperkeratosis, xerosis, ichthyosis,
 CC psoriasis, malignant or benign hyperkeratotic tumours (including those
 CC of the Malpighian mucosa), reactive keratosis, and leucokeratosis (of
 CC the cervix uteri, mouth or during prolapse) can also be treated with
 CC the products.
 SQ Sequence 529 AA;

Query Match 60.3%; Score 35; DB 1; Length 529;
 Best Local Similarity 60.0%; Pred. No. 10;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ADPXXYXPOG 10
 III : III
 Db 514 ADPEVFLPQG 523

RESULT 3

ID W85055 standard; Protein; 529 AA.
 AC W85055;
 DT 05-FEB-1999 (first entry)
 DE Human protein designated corneodesmosin.
 KW Corneodesmosin; intercorneocyte cohesion; skin trophic disorder;
 KW scar formation; epidermis thinning; skin fragility; hyperkeratosis;
 KW xerosis; ichthyosis; psoriasis; hyperkeratotic tumour;
 KW reactive keratosis; and leucokeratosis.
 OS Homo sapiens.
 PN FR2761363-A1.
 PD 02-OCT-1998.
 PF 11-SEP-1997; 011317.
 PR 28-MAR-1997; FR-003899.
 PA (OREA) L'OREAL SA.
 PI Serre G, Simon M, Weber VM;
 DR WPI; 98-586094/50.
 DR N-PSDB; V63541.
 PT New polypeptide, corneodesmosin, from skin and related nucleic acid
 PT - for cosmetic or therapeutic treatment of trophic disorders,
 PT thinning of the epidermis etc., also compositions containing
 PT protease that degrades corneodesmosin
 PS Claim 1; Page 30; 35pp; French.
 CC The present sequence represents a protein designated corneodesmosin.
 CC The protein is involved in intercorneocyte cohesion and is used, in
 CC cosmetic or pharmaceutical compositions, to treat trophic disorders
 CC of the skin and disorders associated with scar formation, and
 CC thinning of the epidermis, particularly the corneal layer, or skin
 CC fragility, to reinforce intercorneocyte cohesion and/or to induce
 CC thickening of the corneal layer. Hyperkeratosis, xerosis, ichthyosis,
 CC psoriasis, malignant or benign hyperkeratotic tumours (including those
 CC of the Malpighian mucosa), reactive keratosis, and leucokeratosis (of
 CC the cervix uteri, mouth or during prolapse) can also be treated with
 CC the products.
 SQ Sequence 529 AA;

Query Match 60.3%; Score 35; DB 1; Length 529;

Best Local Similarity 60.0%; Pred. No. 10;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 ADPXXYXPOG 10
 III : III
 Db 514 ADPEVFLPQG 523

RESULT 4

ID R44997 standard; Protein; 329 AA.
 AC R44997;
 DT 14-JUN-1994 (first entry)
 DE Mob polypeptide encoded by ORF1 of plasmid pBBR1.
 DE high copy number plasmid; broad host range; Rep polypeptide;
 KW Mob polypeptide; mobilisation; Gram negative host.
 KW Bordetella bronchiseptica (strain 587).
 OS Bordetella bronchiseptica
 FH Key Location/Qualifiers
 FT region 1..266
 FT /note= "encoded by complement of nucleotides
 FT 1..799 in Q53521"
 FT region 268..329
 FT /note= "encoded by complement of nucleotides
 FT 2497..2685 in Q53521"
 FT misc_difference 277
 FT /note= "shown as His residue in Figure 1 but as
 FT Asp in Figure 6 and corresponds to GAC
 FT (i.e. Asp) codon in Q53521"
 PN FR2690459-A.
 PD 29-OCT-1993.
 PR 23-APR-1992; 005028.
 PR 23-APR-1992; FR-005028.
 PA (INRM) INST NAT SANTE & RECH MEDICALE.
 PA (INSP) INST PASTEUR.
 PA (INSP) INST PASTEUR LILLE.
 PI Antoine R, Lochet C;
 DR WPI; 93-397119/50.
 DR N-PSDB; Q53521.
 PT New, small plasmid for Bordetella bronchiseptica - encoding only
 PT mobilisation and replication functions, for making expression
 PT vectors with broad host range and high copy number, also derived
 PT from nucleic acid
 PS Claim 19; Fig 1 and Fig 6; 69pp; French.
 CC Plasmid pBBR1 was isolated from B. bronchiseptica strain 587. The
 CC plasmid does not contain any resistance genes, appearing only to
 CC code replication (Rep polypeptide R44998) and mobilisation
 CC (Mob polypeptide R44997) functions. Plasmid pBBR1 is replicable
 CC (at 30-40 copies per cell) in Gram negative hosts e.g. E.coli and
 CC B.pertussis. Heterologous DNA can be inserted at sites which do not
 CC affect the Mob or Rep functions for use as a broad host range
 CC cloning vector.
 SQ Sequence 329 AA;

Query Match 56.9%; Score 33; DB 1; Length 329;
 Best Local Similarity 55.6%; Pred. No. 16;
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPOG 10
 : I I I I I

Db 204 EPRAYAPQG 212

RESULT 5

ID R58479 standard; Protein; 38 AA.
 AC R58479;
 DT 18-APR-1995 (first entry)
 DE TSAR binding domain (SA.9-11) binds streptavidin.
 DE TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; monoclonal antibody;
 KW Sm antigen; mouse model; autoimmune disease; SLE; MAb;

KW systemic lupus erythematosus: streptavidin.
 OS Synthetic.
 FH Key
 FT region
 FT 29. -32
 FT /label- consensus_motif
 PN W09418318-A.
 PD 18-AUG-1994. U00977.
 PF 01-FEB-1994.
 PR 01-FEB-1993; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PR (UUNC-) UNIV NORTH CAROLINA.
 PA Fowles DM, Kay BK;
 DR WPI; 94-279739/34.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Example 7.7; Page 117; 255pp; English.
 CC R58476-90 are amino acid sequences of the binding domain of TSAR
 CC (Totally Synthetic Affinity Reagents) peptides. These particular
 CC peptides are encoded by the streptavidin (SA) binding phase.
 CC The SA TSARs fall into 2 classes. First, the majority of SA binding
 CC peptides share the consensus motif HP(O/M)X (X = a non polar amino
 CC acid), and secondly, there is a minor class of SA binding peptides that
 CC lack any consensus sequence and have no apparent similarity with each
 CC other. The non-variable amino acids at the NH2 and COOH terminals are
 CC not shown. TSAR peptides are generated using generic oligonucleotides
 CC (see 070470-73 for examples). TSARs are concatenated heterofunctional
 CC proteins or peptides, comprising at least two functional regions - a
 CC binding domain with affinity for a ligand and a second effector peptide
 CC portion that is chemically or biologically active. They may further
 CC comprise a linker peptide between the 2 domains. The TSARs or compsns.
 CC comprising a TSAR binding domain can be used in vivo to deliver a
 CC chemically or biologically active moiety, eg. metal ion, radioisotope,
 CC peptide, toxin or enzyme, to the specific target or on the cell. They
 CC can also replace the function of macromolecules eg. monoclonal or
 CC polyclonal antibodies and therefore circumvent the need for complex
 CC methods of hybridoma formation or in vivo antibody production. The TSARs
 CC are easily characterised and have designed activity allowing direct and
 CC rapid detection in a screening process.
 SQ Sequence 38 AA;

Query Match 53.4%; Score 31; DB 1; Length 38;
 Best Local Similarity 55.6%; Pred. No. 4.3;
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPOG 10
 : | | | |
 Db 24 NPFTYHPOG 32

RESULT 6

R58815
 ID R58815 standard; Protein; 243 AA.
 AC R58815;
 DT 13-APR-1995 (first entry)
 DE Human c-myc far upstream element (FUSE) binding protein (FBP)
 DE variant from HL60 clone 3-1
 KW DNA-binding regulator; c-myc; DRONE; FUSE; activator; cis-element;
 KW promoter Pl.
 OS Homo sapiens.
 PN W09419465-A.
 PD 01-SEP-1994.
 PF 22-FEB-1994; U01782.
 PR 22-FEB-1993; US-021608.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Avigan MI, Duncan RC, Levens DL;
 DR N-PSDB; 068911.
 PT New DNA-binding regulator of c-myc expression and its cDNA - used
 PT to develop prods. for diagnosis and therapy of disease states
 PT such as tumour formation

PS Claim 17; Page 48-49; 94pp; English.
 CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far
 CC upstream element) binding protein (FBP) are synonymous. A FUSE which
 CC is required for maximal transcription of c-myc binds a factor (DROME
 CC or FBP). The activator cis-element is approx. 1500 bp 5' of the
 CC human c-myc promoter Pl. A full length FBP cDNA sequence was
 CC assembled from overlapping clones obtd. from cDNA libraries. Source
 CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
 CC BJAB, and PMA/PMA stimulated pooled human peripheral blood
 CC lymphocytes (PBLs). A composite cDNA is given in Q68909 and the
 CC deduced polypeptide sequence, (67.5 kd mol. wt.) is given in R58813.
 CC The clone from an activated PBL cDNA library labeled "3-1" contains an
 CC FBP variant. The AA sequence for the protein encoded by clone
 CC 3-1 can be found in R58815.
 SQ Sequence 243 AA;

Query Match 51.7%; Score 30; DB 1; Length 243;
 Best Local Similarity 62.5%; Pred. No. 50;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PXXYXPOG 10
 : | | | |
 Db 225 PAPPYPOG 232

RESULT 7

R58816
 ID R58816 standard; Protein; 590 AA.
 AC R58816;
 DT 13-APR-1995 (first entry)
 DE Human c-myc far upstream element (FUSE) binding protein (FBP)
 DE variant from PBL clone 31-10.
 KW DNA-binding regulator; c-myc; DRONE; FUSE; activator; cis-element;
 KW Promoter Pl.
 OS Homo sapiens.
 PN W09419465-A.
 PD 01-SEP-1994.
 PF 22-FEB-1994; U01782.
 PR 22-FEB-1993; US-021608.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Avigan MI, Duncan RC, Levens DL;
 DR N-PSDB; 068912.
 PT New DNA-binding regulator of c-myc expression and its cDNA - used
 PT to develop prods. for diagnosis and therapy of disease states
 PT such as tumour formation

PS Claim 20; Page 53-56; 94pp; English.
 CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far
 CC upstream element) binding protein (FBP) are synonymous. A FUSE which
 CC is required for maximal transcription of c-myc binds a factor (DROME
 CC or FBP). The activator cis-element is approx. 1500 bp 5' of the
 CC human c-myc promoter Pl. A full length FBP cDNA sequence was
 CC assembled from overlapping clones obtd. from cDNA libraries. Source
 CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
 CC BJAB, and PMA/PMA stimulated pooled human peripheral blood
 CC lymphocytes (PBLs). A composite cDNA is given in Q68909 and the
 CC deduced polypeptide sequence, (67.5 kd mol. wt.) is given in R58813.
 CC The clone from an activated PBL cDNA library labeled "31-10" contains
 CC an FBP variant. The ORF comprised of bps 135-1991 of Q68909 with
 CC two exceptions. Clone 31-10 contains 63 bps inserted at posn. 238
 CC which probably result from an intron which had not been spliced
 CC out. The inserted bps remain in the ORF. The 31-10 clone also
 CC deviates in that bps 1807-19522 are deleted. This deletion shifts
 CC out of frame the stop codons which would terminate translation in
 CC the other clones. The AA sequence for the protein encoded by clone
 CC 31-10 can be found in R58816.
 SQ Sequence 590 AA;

Query Match 51.7%; Score 30; DB 1; Length 590;
 Best Local Similarity 62.5%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PXXYXPOG 10
| | | | |
Db 486 PAPPYAPQG 493

RESULT 8

ID R58813 standard; Protein; 644 AA.
AC R58813;
DT 13-APR-1995 (first entry)
DE Human c-myc far upstream element (FUSE) binding protein (FBP).
KW DNA-binding regulator; c-myc; DRONE; FUSE; activator; cis-element;
KW promoter P1.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_difference 149
/label= M,I
PN WO9419465-A.
PD 01-SEP-1994.
PF 22-FEB-1994; U01782.
PR 22-FEB-1993; US-021608.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Avigan MI, Duncan RC, Levens DL;
DR N-PSDB; Q68909.
PT New DNA-binding regulator of c-myc expression and its cDNA - used
to develop prods. for diagnosis and therapy of disease states
PT such as tumour formation
PS Claim 5; Page 40-43; 94pp; English.
CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far
upstream element) binding protein (FBP) are synonymous. A FUSE which
is required for maximal transcription of c-myc binds a factor (DROME
or FBP). The activator cis-element is approx. 1500 bp 5' of the
CC human c-myc promoter P1. A full length FBP cDNA sequence was
CC assembled from overlapping clones obt'd. from cDNA libraries. Source
CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
CC BJAB, and PMA/PHA stimulated pooled human peripheral blood
CC lymphocytes (PBLs). A composite cDNA is given in Q68909 and the
CC deduced polypeptide sequence, (67.5 kd mol. wt.) is given in R58813.
SQ Sequence 644 AA;

Query Match 51.7%; Score 30; DB 1; Length 644;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PXXYXPOG 10
| | | | |
Db 502 PAPPYAPQG 509

RESULT 9

ID R58814 standard; Protein; 643 AA.
AC R58814;
DT 13-APR-1995 (first entry)
DE Human c-myc far upstream element (FUSE) binding protein (FBP)
DE variant lacking Ser 97.
KW DNA-binding regulator; c-myc; DRONE; FUSE; activator; cis-element;
KW promoter P1.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_difference 148
/label= M,I
PN WO9419465-A.
PD 01-SEP-1994.
PF 22-FEB-1994; U01782.
PR 22-FEB-1993; US-021608.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Avigan MI, Duncan RC, Levens DL;
DR WPI; 94-294330/36.
DR N-PSDB; Q68910.

PT New DNA-binding regulator of c-myc expression and its cDNA - used
to develop prods. for diagnosis and therapy of disease states
PT such as tumour formation
PS Claim 12; Page 60-63; 94pp; English.
CC DROME (DNA-binding regulator of c-myc expression) and FUSE (far
upstream element) binding protein (FBP) are synonymous. A FUSE which
is required for maximal transcription of c-myc binds a factor (DROME
or FBP). The activator cis-element is approx. 1500 bp 5' of the
CC human c-myc promoter P1. A full length FBP cDNA sequence was
CC assembled from overlapping clones obt'd. from cDNA libraries. Source
CC RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
CC BJAB, and PMA/PHA stimulated pooled human peripheral blood
CC lymphocytes (PBLs). A composite cDNA is given in Q68909 and the
CC deduced polypeptide sequence, (67.5 kd mol. wt.) is given in R58813.
CC Three clones from a BJAB cDNA library and three clones from an
CC activated human PBL cDNA library contain the sequence in Q68909.
CC Three clones from the same BJAB library and two clones from the
CC activated lymphocyte library are lacking bps 316,317 and 318
CC (see Q68910). The mRNA lacking these three bps would encode a
CC variant protein lacking Ser 97 (see R58814).
SQ Sequence 643 AA;

Query Match 51.7%; Score 30; DB 1; Length 643;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PXXYXPOG 10
| | | | |
Db 501 PAPPYAPQG 508

RESULT 10

ID W59912 standard; Protein; 1291 AA.
AC W59912;
DT 20-NOV-1998 (first entry)
DE Amino acid sequence of the mutanase enzyme.
KW Mutanase enzyme; PT box; alpha-1,3 glucoside bond; mutan; plaque;
KW bacteria; teeth.
OS Bacillus sp.
FH Key Location/Qualifiers
FT Peptide 1..31
FT Domain 306..354
/note= "PT box"
PN J10201483-A.
PD 04-AUG-1998.
PF 01-OCT-1997; 284362.
PR 25-NOV-1996; JP-314057.
PA (LLOY) LION CORP.
DR WPI; 98-474495/41.
DR N-PSDB; V53575.
PT Gene encoding a mutanase enzyme - used for prevention and removal
of plaque and bacteria on teeth
PT Disclosure; Pages 11-14; 15pp; Japanese.
CC This is the amino acid sequence of the mutanase enzyme, comprising
CC the a PT box, which decomposes the alpha-1,3 glucoside bond of mutan.
CC The mutanase enzyme is used in the method of the invention for
CC prevention and removal of plaque and bacteria on teeth.
SQ Sequence 1291 AA;

Query Match 51.7%; Score 30; DB 1; Length 1291;
Best Local Similarity 33.3%; Pred. No. 3e+02;
Matches 5; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 ADPXXYXPOGXXXVVG 15
: | | | | | :
Db 928 SDKILYIPEGTFHLG 942

RESULT 11

W72395

ID W72395 standard; Protein; 189 AA.
 AC W72395;
 DE 02-FEB-1999 (first entry)
 DE Pathogen response protein LSD1-interacting protein T.
 KW LSD1-interacting protein T: plant pathogen response; apoptosis;
 KW programmed cell death; disease resistance; herbicide resistance;
 KW transgenic plant; crop protection.
 OS Arabidopsis thaliana.
 PN W0983755-A1.
 PD 03-SEP-1998.
 PF 27-FEB-1998; U04077.
 PR 28-FEB-1997; US-039063.
 PA (UINC-) UNIV NORTH CAROLINA.
 PI Dangl JL, Dietrich RA, Eppl PM, Richberg MH;
 DR WPI; 98-531501/45.
 DR N-PSDB; V66766.
 PT New isolated Arabidopsis genes - useful for producing transgenic
 PT plants which show resistance to cell death caused by pathogens or
 PT herbicides.
 PS Claim 46; Page 68; 88pp; English.
 CC This is the amino acid sequence of LSD1-interacting protein T of
 CC Arabidopsis thaliana. LSD1 interacting genes (see V66755-67) were
 CC isolated from a yeast gene expression library constructed in
 CC plasmid pUG4-5 using RNA from Arabidopsis leaves infected with
 CC Pseudomonas syringae. A two-hybrid system was used with LSD1 short
 CC and long open reading frames (see V66750-51) as bait. LSD1 (see
 CC W72366-67) is a novel polypeptide that regulates the initial
 CC response of plants to pathogens and the subsequent spread of plant
 CC cell death engendered by infection. Since the inactivation of
 CC LSD1 by mutation leads to enhanced disease resistance, LSD1
 CC partner proteins represent novel targets for engineering plants
 CC with enhanced resistance to pathogens. Thus, the invention
 CC includes all proteins (see W72384-96) that interact with the cell
 CC death regulator LSD1.
 CC Sequence 189 AA;
 SQ

Query Match 51.7%; Score 30; DB 1; Length 189;
 Best Local Similarity 38.5%; Pred. No. 38;
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 PXXYXPGQXXVVG 15
 Db 142 PYGLPFGTAPIG 154

RESULT 12

R31955
 ID R31955 standard; Protein; 441 AA.
 AC R31955;
 DT 06-JUN-1993 (first entry)
 DE Sequence encoded by glycoprotein G gene.
 DE IBR glycoprotein E gene; unique short 2 gene.
 OS Infectious bovine rhinotracheitis.
 PN W09302104-A.
 PD 04-FEB-1993.
 PF 20-JUL-1992; U06034.
 PR 18-JUL-1991; US-732584.
 PA (SVTR) SYNTRON CORP.
 PI Cochran MD, Macdonald RD;
 DR WPI; 93-058725/07.
 DR N-PSDB; Q36768.
 PT Recombinant infectious bovine rhinotracheitis virus - provides
 PT isolated DNA encoding gpE glyco:protein, gpG glyco:protein and
 PT unique short 2 genes of the virus
 PS Example; Fig 8; 240pp; English.
 CC The sequence of approximately 1400 base pairs of the HindIII K
 CC fragment, starting approximately 2800 base pairs downstream of the
 CC HindIII K/HindIII O junction, are shown. The glycoprotein G (gpG)
 CC gene is transcribed away from the HindIII K/HindIII O junction.
 CC Sequence 441 AA;
 SQ

Query Match 50.0%; Score 29; DB 1; Length 441;
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;
 Matches 5; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADPXXYXPGQ 10
 Db 247 ADPIDYADEG 256

RESULT 13

R85475
 ID R85475 standard; Protein; 161 AA.
 AC R85475;
 DT 12-MAR-1996 (first entry)
 DE PTLV-L Rex protein.
 KW PTLV-L; STL; HTLV; baboon; vaccine; diagnostic; PH969;
 KW Rex protein.
 OS T-lymphotropic primate virus.
 PN W09526405-A1.
 PD 05-OCT-1995.
 PR 23-MAR-1995; BE0026.
 PR 25-MAR-1994; EP-200804.
 PR 08-JUN-1994; NL-000932.
 PA (REGA-) STICHTING REGA VZW.
 PI Desmyter JHMA, Goubau PFJ;
 DR WPI; 95-351325/45.
 DR N-PSDB; T05713.
 DT New T-lymphotropic primate virus PTLV-L isolated from baboon(s) -
 PT homologous to human viruses, also related polypeptide(s), nucleic
 PT acid, antibodies etc., useful in diagnostic assays and vaccines
 PS Example 2; Fig 5-1 to 5-8; 42pp; English.
 CC The Rex protein (R85475) is the product of the rex gene (see
 CC T05713) of a novel T-lymphotropic primate virus, PTLV-L.
 CC isolated from an Eritrean wild baboon (Papio hamadryas). Rex and
 CC Tax (R85474) of PTLV-L are coded by a messenger generated by
 CC splicing on sd-tr and sa-pX3. Epitopes of PTLV-L can be used in
 CC vaccine prodn. or to raise diagnostic antibodies.
 CC Sequence 161 AA;
 SQ

Query Match 50.0%; Score 29; DB 1; Length 161;
 Best Local Similarity 55.6%; Pred. No. 52;
 Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 DPXXYXPGQ 10
 Db 138 DPXYTPCG 146

RESULT 14

R77522
 ID R77522 standard; Protein; 516 AA.
 AC R77522;
 DT 20-MAR-1996 (first entry)
 DE Cytochrome P450 CS.
 KW Cytochrome P450; P450-monooxygenase; insecticide; housefly.
 OS Musca domestica CS.
 PN W09530745-A1.
 PD 16-NOV-1995.
 PF 08-MAY-1995; U05758.
 PR 10-MAY-1994; US-241388.
 PA (CORR) CORNELL RES FOUND INC.
 PI Scott JG, Tomita T;
 DR WPI; 95-404112/51.
 DR N-PSDB; T05526.
 PT DNA encoding cytochrome P450-lpr - used for insect control,
 PT bioremediation of insecticides or reducing crop sensitivity to
 PT pesticides
 PS Example 11; Page 39-40; 87pp; English.
 CC The amino acid sequence of cytochrome P450-cs (R77522) isolated
 CC from insecticide-susceptible adult housefly strain CS showed only
 CC 2 differences from that of P450-lpr, a cytochrome capable of
 CC metabolising insecticides, obtd. from housefly strain Learn-PyR.
 CC

CC Ile-220 and Ile-469 in P450-lpr are replaced by Phe and Met,
 CC respectively, in CS.
 SQ Sequence 516 AA;

Query Match 50.0%; Score 29; DB 1; Length 516;
 Best Local Similarity 55.6%; Pred. No. 1.8e+02;
 Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 DPXXYXPOG 10
 |||||
 Db 447 DPAAYMPEG 455

RESULT 15

R77511
 ID R77511 standard; Protein; 516 AA.
 AC R77511:
 DT 20-MAR-1996 (first entry)
 DE Cytochrome-P450-lpr.
 KW Cytochrome P450-lpr; P450-monoxygenase; housefly; insecticide;
 KW biological control; pesticide degradation; bioremediation;
 KW transgenic plant; crop improvement; insect resistance.
 OS Musca domestica Learn-Pyr.
 FH Key Location/Qualifiers
 FT peptide 1..19
 FT /label= Sig-peptide
 FT /note= "amino acids 1-19 constitute a
 membrane-anchor signal"
 FT 20..516
 FT /label= Mat_protein
 PN W09530745-A1.
 PD 16-NOV-1995.
 PF 08-MAY-1995; U05758
 PR 10-MAY-1994; US-241388.
 PA (CORR) CORNELL RES FOUND INC.
 PI Scott JG, Tomita T;
 DR WPI; 95-404112/51.
 DR N-PSDB; T05517.
 PT DNA encoding cytochrome P450-lpr - used for insect control,
 PT bioremediation of insecticides or reducing crop sensitivity to
 PT pesticides
 PS Claim 5; Page 44-46; 87pp; English.
 CC Cytochrome P450-lpr (R77511) Obtd. from adult housefly strain
 CC Learn-Pyr, selected with permethrin for 22 generations, is capable of
 CC metabolising pyrethroids and activating organophosphate insecticides.
 CC Recombinant P450-lpr, produced by expression of encoding cDNA (T05517),
 CC can be used for insecticide bioremediation. Expression in e.g. a
 CC baculovirus vector allows biological control of larval and adult insect
 CC pests, while expression in transgenic plants reduces crop sensitivity
 CC to pesticides.
 SQ Sequence 516 AA;

Query Match 50.0%; Score 29; DB 1; Length 516;
 Best Local Similarity 55.6%; Pred. No. 1.8e+02;
 Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 DPXXYXPOG 10
 |||||
 Db 447 DPAAYMPEG 455

Search completed: November 13, 1999, 18:59:59
 Job time: 138 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:33 ; Search time 64.87 Seconds
(without alignments)
2.643 Million cell updates/sec

Title: US-08-913-430-11

Perfect score: 58

Sequence: 1 ADPXXYPGCGXXVG 15

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/PCTUS9_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	53.4	38	1	US-08-176-500-106
2	31	53.4	38	1	US-08-471-052A-106
3	31	53.4	38	1	US-08-189-331-106
4	31	53.4	38	2	US-08-471-939-106
5	31	53.4	38	2	US-08-471-800-106
6	31	53.4	38	2	US-08-471-068-106
7	30	51.7	644	1	US-08-021-608D-2
8	30	51.7	243	1	US-08-021-608D-6
9	30	51.7	590	1	US-08-021-608D-8
10	30	51.7	643	1	US-08-021-608D-10
11	30	51.7	644	1	US-08-726-160-2
12	30	51.7	243	1	US-08-726-160-6
13	30	51.7	590	1	US-08-726-160-8
14	30	51.7	643	1	US-08-726-160-10
15	30	51.7	644	3	PCT-US94-01782-2
16	30	51.7	243	3	PCT-US94-01782-6
17	30	51.7	590	3	PCT-US94-01782-8
18	30	51.7	643	3	PCT-US94-01782-10
19	29	50.0	517	1	US-08-457-274A-2
20	29	50.0	517	1	US-08-457-274A-28
21	29	50.0	441	2	US-08-191-866D-21
22	29	50.0	36	2	US-08-340-539A-21
23	29	50.0	441	2	US-08-185-949B-21
24	29	50.0	748	2	US-08-920-234-2
25	29	50.0	748	2	US-08-937-931-4
26	29	50.0	749	2	US-08-937-931-8
27	29	50.0	517	3	PCT-US95-05758-2
28	29	50.0	517	3	PCT-US95-05758-28
29	28	48.3	2396	1	US-08-157-005-2
30	28	48.3	1308	2	US-08-484-438-2
31	28	48.3	541	2	US-08-484-438-6
32	28	48.3	705	2	US-08-456-647B-4
33	28	48.3	705	2	US-08-237-401A-4
34	27	46.6	276	1	US-07-952-817-27
35	27	46.6	1068	1	US-08-396-479B-12
36	27	46.6	742	1	US-08-347-718B-1
37	27	46.6	742	1	US-08-347-718B-2
38	27	46.6	1068	1	US-08-818-823-12
39	27	46.6	504	1	US-08-457-274A-25

40 27 46.6 745 2 US-08-445-050-2 Sequence 2, Appli
41 27 46.6 722 2 US-08-445-050-3 Sequence 3, Appli
42 27 46.6 535 2 US-08-445-050-4 Sequence 4, Appli
43 27 46.6 546 2 US-08-445-050-5 Sequence 5, Appli
44 27 46.6 568 2 US-08-445-050-6 Sequence 6, Appli
45 27 46.6 722 2 US-08-445-050-7 Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-176-500-106
; Sequence 106, Application US/08176500
; Patent No. 5498538
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 141
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/176,500
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/013,416
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-143
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-176-500-106

Query Match 53.4%; Score 31; DB 1; Length 38;
Best Local Similarity 55.6%; Pred. NO. 2.7;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYPQG 10

Db 24 NPFTYHPQG 32

RESULT 2

US-08-471-052A-106
; Sequence 106, Application US/08471052A
; Patent No. 5625033
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.

;
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 166
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,052A
; FILING DATE: 06-JUNE-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-471-052A-106

Query Match 53.4%; Score 31; DB 1; Length 38;
Best Local Similarity 55.6%; Pred. No. 2.7;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXXPQG 10
; : | | | |
Db 24 NPFTYHPQG 32

RESULT 3
US-08-189-331-106
; Sequence 106, Application US/08189331
; Patent No. 5747334
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 186
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/189,331
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872

;
; REFERENCE/DOCKET NUMBER: 1101-155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEX: 212 869-8864/9741
; TELEFAX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-189-331-106

Query Match 53.4%; Score 31; DB 1; Length 38;
Best Local Similarity 55.6%; Pred. No. 2.7;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXXPQG 10
; : | | | |
Db 24 NPFTYHPQG 32

RESULT 4
US-08-471-939-106
; Sequence 106, Application US/08471939
; Patent No. 5844076
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 141
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,939
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/013,416
; FILING DATE: 01-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-143
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEX: 212 869-8864/9741
; TELEFAX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-471-939-106

Query Match 53.4%; Score 31; DB 2; Length 38;
Best Local Similarity 55.6%; Pred. No. 2.7;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYPQG 10
: | | | |
Db 24 NPFTYHPQG 32

RESULT 5
US-08-471-800-106
; Sequence 106, Application US/08471800
; Patent No. 5852167
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 141
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471.800
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/013,416
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 19,872
; REFERENCE/DOCKET NUMBER: 1101-143
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-471-800-106

Query Match 53.4%; Score 31; DB 2; Length 38;
Best Local Similarity 55.6%; Pred. No. 2.7;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYPQG 10
: | | | |
Db 24 NPFTYHPQG 32

RESULT 6
US-08-471-068-106
; Sequence 106, Application US/08471068
; Patent No. 5948635
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 186
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas

; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/189,331
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-471-068-106

Query Match 53.4%; Score 31; DB 2; Length 38;
Best Local Similarity 55.6%; Pred. No. 2.7;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYPQG 10
: | | | |
Db 24 NPFTYHPQG 32

RESULT 7
US-08-021-608D-2
; Sequence 2, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,608D
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 644
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Human
CELL LINE: HL60
FEATURE:
OTHER INFORMATION:
OTHER INFORMATION: Amino Acid 149 (Xaa) is Met or Ile

US-08-021-608D-2

Query Match 51.7%; Score 30; DB 1; Length 644;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PXXYXPG 10
| | | | |
Db 502 PAPPYAPQG 509

RESULT 8
US-08-021-608D-6
Sequence 6, Application US/08021608D
Patent No. 5580760
GENERAL INFORMATION:
APPLICANT: LEVENS, DAVID L., DUNCAN,
APPLICANT: ROBERT C., AND AVIGAN, MARK I.
TITLE OF INVENTION: NOVEL FUSE BINDING
TITLE OF INVENTION: NOVEL FUSE BINDING
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,608D
FILING DATE: 22-FEB-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 243

TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Human
CELL TYPE: HL60
US-08-021-608D-6

Query Match 51.7%; Score 30; DB 1; Length 243;
Best Local Similarity 62.5%; Pred. No. 32;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PXXYXPG 10
| | | | |
Db 225 PAPPYAPQG 232

RESULT 9
US-08-021-608D-8
Sequence 8, Application US/08021608D
Patent No. 5580760
GENERAL INFORMATION:
APPLICANT: LEVENS, DAVID L., DUNCAN,
APPLICANT: ROBERT C., AND AVIGAN, MARK I.
TITLE OF INVENTION: NOVEL FUSE BINDING
TITLE OF INVENTION: NOVEL FUSE BINDING
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,608D
FILING DATE: 22-FEB-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 590
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Human
CELL LINE: HL60
US-08-021-608D-8

Query Match 51.7%; Score 30; DB 1; Length 590;
Best Local Similarity 62.5%; Pred. No. 84;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PXXYXPOG 10
| | | | |
Db 486 PAPIAPOG 493

RESULT 10
US-08-021-608D-10
; Sequence 10, Application US/08021608D
; Patent No. 5580760
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021.608D
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide/Protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION:
; Amino Acid 148 (Xaa) is Met or Ile

US-08-021-608D-10

Query Match 51.7%; Score 30; DB 1; Length 643;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PXXYXPOG 10
| | | | |
Db 501 PAPIAPOG 508

RESULT 11
US-08-726-160-2
; Sequence 2, Application US/08726160

; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/021,608
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 644
; TYPE: Amino Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide/Protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Human
; CELL LINE: HL60
; FEATURE:
; OTHER INFORMATION:
; Amino Acid 149 (Xaa) is Met or Ile

US-08-726-160-2

Query Match 51.7%; Score 30; DB 1; Length 644;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PXXYXPOG 10
| | | | |
Db 502 PAPIAPOG 509

RESULT 12
US-08-726-160-6
; Sequence 6, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE

CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,160
FILING DATE: 04-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/021,608
FILING DATE: 22-FEB-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063US1
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 243
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Human
CELL TYPE: HL60
US-08-726-160-6

Query Match 51.7%; Score 30; DB 1; Length 243;
Best Local Similarity 62.5%; Pred. No. 32;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PXXYXPQG 10
| | | | |
Db 225 PAPAPOG 232

RESULT 13
US-08-726-160-8
; Sequence 8, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: NOVEL FUSE BINDING
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/021,608
FILING DATE: 22-FEB-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063US1
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 590
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Human
CELL LINE: HL60
US-08-726-160-8

Query Match 51.7%; Score 30; DB 1; Length 590;
Best Local Similarity 62.5%; Pred. No. 84;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PXXYXPQG 10
| | | | |
Db 486 PAPAPOG 493

RESULT 14
US-08-726-160-10
; Sequence 10, Application US/08726160
; Patent No. 5734016
; GENERAL INFORMATION:
; APPLICANT: LEVENS, DAVID L., DUNCAN,
; APPLICANT: ROBERT C., AND AVIGAN, MARK I.
; TITLE OF INVENTION: NOVEL FUSE BINDING
; TITLE OF INVENTION: NOVEL FUSE BINDING
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,160
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/021,608
; FILING DATE: 22-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4063US1
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:

LENGTH: 643
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: No
ORIGINAL SOURCE:
ORGANISM: Human
CELL LINE: HL60
FEATURE:
OTHER INFORMATION: Amino acid 148 (Xaa) is Met or Ile
US-08-726-160-10

Query Match 51.7%; Score 30; DB 1; Length 643;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 3 PXXYXPQG 10
Db 501 PARYAPQG 508

RESULT 15
PCT-US94-01782-2
Sequence 2, Application PC/TUS9401782
GENERAL INFORMATION:
APPLICANT: THE GOVERNMENT OF THE UNITED STATES
APPLICANT: AS REPRESENTED BY THE SECRETARY, DEPARTMENT OF
APPLICANT: HEALTH AND HUMAN SERVICES
TITLE OF INVENTION: NOVEL FUSE BINDING
TITLE OF INVENTION: PROTEIN AND CDNA THEREFOR
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01782
FILING DATE: 22-FEB-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/021,608
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4063PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-8849
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 644
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Peptide/Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Human
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:

HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE: HL60
ORGANELLE:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
OTHER INFORMATION: Amino Acid 149 (Xaa) is Met or Ile
PCT-US94-01782-2

Query Match 51.7%; Score 30; DB 3; Length 644;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 3 PXXYXPQG 10
Db 502 PARYAPQG 509

Search completed: November 13, 1999, 10:56:34
Job time: 1363 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:35 ; Search time 139.86 Seconds
(without alignments)
6.601 Million cell updates/sec

Title: US-08-913-430-11

Perfect score: 58
Sequence: 1 ADPXXYXPQXXXXVG 15

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPTREMBL_10:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	60.3	530	2	O34022 chlamydia p
2	35	60.3	529	4	O43509 homo sapien
3	34	58.6	352	5	O18015 caenorhabdi
4	33	56.9	339	2	O54530 escherichia
5	33	56.9	329	2	O925R6 bordetella
6	32	55.2	534	2	O87009 burkholderi
7	32	55.2	350	5	O25413 lyndae sta
8	32	55.2	858	5	O17647 caenorhabdi
9	32	55.2	419	10	O49429 arbidopsis
10	31	53.4	295	1	O26704 methanobact
11	31	53.4	858	2	O69213 anabaena sp
12	31	53.4	728	2	O92512 streptomyce
13	31	53.4	706	4	O15024 homo sapien
14	31	53.4	104	5	P92184 drosophila
15	31	53.4	104	5	P92192 drosophila
16	31	53.4	100	5	P92194 drosophila
17	31	53.4	104	5	O96905 drosophila
18	31	53.4	123	10	O04644 arbidopsis
19	31	53.4	749	11	O63643 rattus norv
20	31	53.4	114	13	P79900 oncorhynchu
21	30	51.7	868	1	O29867 archaeoglob
22	30	51.7	182	2	P97062 streptomyce
23	30	51.7	885	2	O55683 synechocyst
24	30	51.7	1336	3	P78966 schizosacch
25	30	51.7	855	3	O42670 schizosacch
26	30	51.7	644	4	O12828 homo sapien
27	30	51.7	533	5	O17252 caenorhabdi
28	30	51.7	208	10	O04033 arbidopsis
29	30	51.7	617	10	O65778 cynara scol

ALIGNMENTS

RESULT 1

O34022
ID O34022 PRELIMINARY; PRT; 530 AA.
AC O34022;
DT 01-JAN-1998 (TREMREL. 05, Created)
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
DE SIMILAR TO MALQ.
OS Chlamydia psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GUINEA PIG INCLUSION CONJUNCTIVITIS;
RX MEDLINE; 97426043.
RA HSIA R.C., PANNEKOEK Y., INGEROWSKI E., BAVOIL P.M.;
RT "Type III secretion genes identify a putative virulence locus of Chlamydia.";
RL Mol. Microbiol. 25:351-359(1997).
DR EMBL; U88070; AAB71512.1;
SQ SEQUENCE 530 AA; 61326 MW; 4EB1F2C CRC32;

Query Match 60.3%; Score 35; DB 2; Length 530;

Best Local Similarity 66.7%; Pred No. 10;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPQG 10

|||
|||
Db 337 DPXXYXPQG 345

RESULT 2

O43509
ID O43509 PRELIMINARY; PRT; 529 AA.
AC O43509;
DT 01-JUN-1998 (TREMREL. 06, Created)
DT 01-JUN-1998 (TREMREL. 06, Last sequence update)
DT 01-AUG-1998 (TREMREL. 07, Last annotation update)
DE CORNEODESMOSIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EPIDERMIS;
RA GUERRIN M., SIMON M., MONTEZIN M., VINCENT C.H., SERRE G.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF030130; AAC24196.1;
SQ SEQUENCE 529 AA; 51479 MW; 76E7D22D CRC32;

Query Match 60.3%; Score 35; DB 4; Length 529;

Best Local Similarity 60.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ADPXXYXPOG 10
||| : |||
Db 514 ADPEVFLPOG 523

RESULT 3

O18015
ID O18015 PRELIMINARY; PRT; 352 AA.
AC O18015;
DT 01-JAN-1998 (TREMREL. 05, Created)
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
DT 01-JAN-1999 (TREMREL. 09, Last annotation update)
DE T02E9.2 PROTEIN.
GN T02E9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA KERSHAW J.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Nature 368:32-38(1994).
RL EMBL: 293387; CAB07652.1; -
SQ SEQUENCE 352 AA; 36430 MW; E5C3FD1 CRC32;

Query Match 58.6%; Score 34; DB 5; Length 352;
Best Local Similarity 46.7%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ADPXXYXPOGXXXVG 15
||| : |||
Db 83 AAPAPYQPGGYQAG 97

RESULT 4

O54530
ID O54530 PRELIMINARY; PRT; 339 AA.
AC O54530;
DT 01-JUN-1998 (TREMREL. 06, Created)
DT 01-JUN-1998 (TREMREL. 06, Last sequence update)
DT 01-MAY-1999 (TREMREL. 10, Last annotation update)
DE MOBILIZATION PROTEIN (FRAGMENT).
OS Escherichia coli.
OG Plasmid pBHRK19, plasmid pBHRK18, and plasmid pBHR1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98172738.
RA GABANT P., SZPIRER C.Y., COUTURIER M., FAELAN M.;
RT "Direct selection cloning vectors adapted to the genetic analysis of
RT gram-negative bacteria and their plasmids";
RL Gene 207:87-92(1998).
RN [2]

RP SEQUENCE FROM N.A.
RA SZPIRER C.Y., GABANT P., COUTURIER M., FAELAN M.;
RT "Study of the mobilization and maintenance elements of the pBHR1
RT broad host range cloning vector";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y14437; CAA74780.1; -
DR EMBL: Y14438; CAA74784.1; -
DR EMBL: Y14439; CAA74788.1; -
DR PFAM; PF01076; Mob_Pre; 1.
KW Plasmid.
FT NON_TER 339 339
SQ SEQUENCE 339 AA; 37834 MW; 212E86E2 CRC32;

Query Match 56.9%; Score 33; DB 2; Length 339;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPOG 10
:| | |||
Db 204 EPRAYAPQG 212

RESULT 5

Q925R6
ID Q925R6 PRELIMINARY; PRT; 329 AA.
AC Q925R6;
DT 01-MAY-1999 (TREMREL. 10, Created)
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
DT 01-MAY-1999 (TREMREL. 10, Last annotation update)
GN MOB PROTEIN.
OS MOB.
OS Bordetella bronchiseptica.
OG Plasmid pBHR1.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S87;
RX MEDLINE: 92334155.
RA ANTOINE R., LOCHT C.;
RT "Isolation and molecular characterization of a novel broad-host-range
RT plasmid from Bordetella bronchiseptica with sequence similarities to
RT plasmids from gram-positive organisms";
RL Mol. Microbiol. 6:1785-1799(1992).
DR EMBL: X66730; CAA747289.2; -
KW Plasmid.
SQ SEQUENCE 329 AA; 36707 MW; 791810B7 CRC32;

Query Match 56.9%; Score 33; DB 2; Length 329;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPOG 10
:| | |||
Db 204 EPRAYAPQG 212

RESULT 6

O87009
ID O87009 PRELIMINARY; PRT; 534 AA.
AC O87009;
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
DE CHLOROPHENOL-4-MONOOXYGENASE COMPONENT 2.
GN TFTD.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=AC1100;
 RX MEDLINE; 98268944.
 RA HUBNER A., DANGANAN C.E., XUN L., CHAKRABARTY A.M., HENDRICKSON W.;
 RT "Genes for 2,4,5-trichlorophenoxyacetic acid metabolism in
 Burkholderia cepacia AC1100: characterization of the tftC and tftD
 genes and locations of the tft operons on multiple replicons.";
 RL Appl. Environ. Microbiol. 64:2086-2093(1998).
 DR EMBL; U83405; AAC23548.1; -.
 KW Monoxygenase.
 SQ SEQUENCE 534 AA; 60273 MW; A086A820 CRC32;

Query Match 55.2%; Score 32; DB 2; Length 534;
 Best Local Similarity 55.6%; Pred. No. 44;
 Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ADPXXYXPQ 9
 Db 487 AEPTAYAPQ 495

RESULT 7
 Q25413 PRELIMINARY; PRT; 350 AA.
 ID Q25413; Q25406;
 AC Q25413; Q25406;
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
 DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
 DE MYOMODULIN NEUROPEPTIDE PRECURSOR.
 OS Lymnaea stagnalis (Great pond snail).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
 OC Lymnaeidae; Lymnaea.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA KELLET E., PERRY S.J., SANTANA N., WORSTER B.M., BENJAMIN P.R.,
 RA BURKE J.F.;
 RL J. Neurosci. 0:0-0(0).
 DR EMBL; X96933; CAA65635.1; -.
 DR EMBL; X96934; CAA65636.1; -.
 KW Signal; Neuropeptide.
 FT SIGNAL
 FT CHAIN 21 350
 SQ SEQUENCE 350 AA; 40388 MW; 54A42CE6 CRC32;

Query Match 55.2%; Score 32; DB 5; Length 350;
 Best Local Similarity 40.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 ADPXXYXPQXXXVG 15
 Db 148 AEPYLPDGFYVG 162

RESULT 8
 Q17647 PRELIMINARY; PRT; 858 AA.
 ID Q17647; Q2025;
 AC Q17647; Q2025;
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
 DT 01-MAY-1999 (TREMREL. 10, Last annotation update)
 DE C05C10.6 PROTEIN.
 GN C05C10.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MATTHEWS P.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z48178; CAA88206.1; -.
 DR EMBL; Z66515; CAA88206.1; JOINED.

DR EMBL; Z66515; CAA91354.1; -.
 DR EMBL; Z48178; CAA91354.1; JOINED.
 DR PFAM; PF00400; WD40; 4.
 SQ SEQUENCE 858 AA; 93639 MW; 2682BABC CRC32;

Query Match 55.2%; Score 32; DB 5; Length 858;
 Best Local Similarity 46.2%; Pred. No. 73;
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 PXXYXPQXXXVG 15
 Db 506 PGRYVPQSSNSG 518

RESULT 9
 O49429 PRELIMINARY; PRT; 419 AA.
 ID O49429;
 AC O49429;
 DT 01-JUN-1998 (TREMREL. 06, Created)
 DT 01-JUN-1998 (TREMREL. 06, Last sequence update)
 DT 01-JUN-1998 (TREMREL. 06, Last annotation update)
 DE DAG-LIKE PROTEIN.
 GN F18F4.120.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., MURPHY G., DROST L., HALL C., HUDSON S., RIDLEY P.,
 RA BANCROFT I., MEWES H.W., MAYER K., SCHUELLER C.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL021637; CAA16610.1; -.
 SQ SEQUENCE 419 AA; 45157 MW; 6C7BE4DE CRC32;

Query Match 55.2%; Score 32; DB 10; Length 419;
 Best Local Similarity 46.2%; Pred. No. 34;
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 PXXYXPQXXXVG 15
 Db 330 PRNYPQAGNFG 342

RESULT 10
 O26704 PRELIMINARY; PRT; 295 AA.
 ID O26704;
 AC O26704;
 DT 01-JAN-1998 (TREMREL. 05, Created)
 DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
 DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
 DE ADHESION PROTEIN.
 GN MTH604.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RX MEDLINE; 98037514.
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
 RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
 RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,
 RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
 RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
 RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,

RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000842; AA85110.1; -.
DR PFAM; PF01297; Lipoprotein_4; 1.
SQ SEQUENCE 295 AA; 33074 MW; 244FBBE9 CRC32;

Query Match 53.4%; Score 31; DB 1; Length 295;
Best Local Similarity 55.6%; Pred. No. 37; Indels 0; Gaps 0;
Matches 5; Conservative 1; Mismatches 3;

QY 1 ADPXXYXPQ 9
||| | | |
Db 67 ADPHTYPEPE 75

RESULT 11
O69213 PRELIMINARY; PRT; 858 AA.
AC O69213;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE PROTEIN SERINE-THREONINE PHOSPHATASE, PRPA.
GN PRPA.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-PCC 7120;
RX MEDLINE; 98241521.
RA ZHANG C.C., FRURY A., PENG L.;
RT "Molecular and genetic analysis of two closely linked genes that
encode, respectively, a protein phosphatase 1/2A/2B homolog and a
protein kinase homolog in the cyanobacterium Anabaena sp. strain PCC
7120.";
RL J. Bacteriol. 180:2616-2622(1998).
DR EMBL; AJ224354; CAAL191.1; -.
SQ SEQUENCE 858 AA; 96869 MW; 109A4F20 CRC32;

Query Match 53.4%; Score 31; DB 2; Length 858;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 PXYXPQGXV 14
| | | | | |
Db 218 PYYHPQGRKAV 229

RESULT 12
Q92512 PRELIMINARY; PRT; 728 AA.
AC Q92512;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ABC EXCISION NUCLEASE SUBUNIT C.
GN SCC54.13C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA SEEGER K., HARRIS D.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RC SEQUENCE FROM N.A.
RA STRAIN-A3(2);
RA BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RC SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL035591; CAB38143.1; -.
SQ SEQUENCE 728 AA; 80841 MW; A9285AF0 CRC32;

Query Match 53.4%; Score 31; DB 2; Length 728;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ADPXXYXPQ 9
||| | | |
Db 2 ADPSSYRPR 10

RESULT 13
O15024 PRELIMINARY; PRT; 706 AA.
AC O15024;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE KIAA0307.
GN KIAA0307.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 97349984.
RA NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002305; BAA20766.1; -.
DR PFAM; PF00010; HLH; 1.
DR PFAM; PF00989; PAS; 1.
SQ SEQUENCE 706 AA; 77614 MW; 54D86E69 CRC32;

Query Match 53.4%; Score 31; DB 4; Length 706;
Best Local Similarity 62.5%; Pred. No. 96;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ADPXXYXP 8
||| | | |
Db 597 ADPSSYSP 604

RESULT 14
P92184 PRELIMINARY; PRT; 104 AA.
AC P92184;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE PUTATIVE CUTICLE PROTEIN LCP6.
GN LCP6.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC Drosophilidae; Drosophila.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-CANTON S;
RA CHARLES J.P., CHIHARA C., NEJAD S., RIDDIFORD L.M.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U84757; AAB48464.1; -;
DR EMBL; U84756; AAB48463.1; -;
DR FLYBASE; FBgn002537; Lcp6.
DR PFAM; PF00379; insect_cuticle; 1.
SQ SEQUENCE 104 AA; 11279 MW; 4446FF46 CRC32;

Query Match 53.4%; Score 31; DB 5; Length 104;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ADPXXYPQG 10
|||
Db 86 ADENGYQPQG 95

RESULT 15

P92192
ID P92192 PRELIMINARY; PRT; 104 AA.
AC P92192;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03; Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10; Last annotation update)
DE CUTICLE PROTEIN DCP3 BETA.
GN DCP3-BETA OR DCP3-ALPHA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
OC Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISO-1;
RA CHARLES J.P., CHIHARA C., NEJAD S., RIDDIFORD L.M.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U84747; AAB88066.1; -;
DR EMBL; U84746; AAB88065.1; -;
DR FLYBASE; FBgn0020643; Lcp65Ab2.
DR FLYBASE; FBgn0020644; Lcp65Ab1.
DR PFAM; PF00379; insect_cuticle; 1.
SQ SEQUENCE 104 AA; 11267 MW; 80FD8C20 CRC32;

Query Match 53.4%; Score 31; DB 5; Length 104;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ADPXXYPQG 10
|||
Db 86 ADENGYQPQG 95

Search completed: November 13, 1999, 12:55:37
Job time: 3036 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:38 ; Search time 251.81 seconds
(without alignments)
3.771 Million cell updates/sec

Title: US-08-913-430-11

Perfect score: 58

Sequence: 1 ADPXXYPQGXVVG 15

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database :

Pending_Patents_AA.*

1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep.*

2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*

3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*

4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*

5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*

6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*

7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep.*

8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep.*

9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*

10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*

11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*

12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*

13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*

14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*

15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*

16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*

17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*

18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*

19: /cgn2_6/ptodata/1/paa/US060_COMB.pep.*

20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*

21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*

22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*

23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*

24: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	79.3	15	13	US-08-913-430-11
2	35	60.3	529	16	US-09-206-817-3
3	34	58.6	479	23	US-09-417-507-38191
4	31	53.4	38	8	US-08-471-052-106
5	31	53.4	38	8	US-08-471-927-106
6	31	53.4	446	10	US-08-660-451A-12
7	31	53.4	338	16	US-09-248-796-18170
8	31	53.4	137	16	US-09-270-767-41428
9	31	53.4	261	17	US-09-328-352-6263
10	31	53.4	338	19	US-08-096-409-18170
11	31	53.4	121	19	US-60-138-830-326
12	31	53.4	249	19	US-60-140-804-711
13	31	53.4	60	19	US-60-140-956-2074
14	31	53.4	79	19	US-60-140-956-2309
15	31	53.4	393	19	US-60-143-753-271
16	31	53.4	657	19	US-60-145-050-183
17	31	53.4	112	19	US-60-145-138-513
18	31	53.4	79	19	US-60-145-138-915
19	31	53.4	79	19	US-60-146-055-650

Sequence 753, App
Sequence 37221, A
Sequence 89, Appl
Sequence 367, App
Sequence 367, App
Sequence 305, App
Sequence 89, Appl
Sequence 367, App
Sequence 20418, A
Sequence 43830, A
Sequence 5107, Ap
Sequence 6328, Ap
Sequence 20418, A
Sequence 526, App
Sequence 28947, A
Sequence 29168, A
Sequence 38625, A
Sequence 4, Appl
Sequence 8, Appl
Sequence 13, Appl
Sequence 2, Appl
Sequence 28, Appl
Sequence 21, Appl
Sequence 7, Appl
Sequence 10, Appl
Sequence 10, Appl

US-60-146-055-753
US-09-417-507-37221
PCT-US98-04077-89
PCT-US98-25247-367
PCT-US98-25247-367
PCT-US99-13418-305
US-09-032-485A-89
US-09-199-637-367
US-09-248-796-20418
US-09-270-767-43830
US-09-328-352-5107
US-09-328-352-6328
US-60-096-409-20418
US-60-138-676-526
US-09-417-507-28947
US-09-417-507-29168
US-09-417-507-38625
PCT-US97-15099-4
PCT-US97-15099-8
PCT-US99-15817-13
US-08-241-388-2
US-08-241-388-28
US-08-379-647A-21
US-08-718-318-7
US-08-960-022-10
US-09-064-703-10

ALIGNMENTS

RESULT 1
US-08-913-430-11
; Sequence 11, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 11
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (4)
; OTHER INFORMATION: Residue may be Phe or Ile
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (5)
; OTHER INFORMATION: Residue may be Arg or Glu
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (7)
; OTHER INFORMATION: Residue may be Val or Ala
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (11)
; OTHER INFORMATION: Residue may be Gln or Ala
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: (12)
; OTHER INFORMATION: Undetermined
; FEATURE:

QY 2 DPXXYXPQGXXVVG 15
||| |||
Db 142 DPEGYEPEGLAGSG 155

```

RESULT 5
US-08-471-927-106
: Sequence 106, Application US/08471927
:
: GENERAL INFORMATION:
:
: APPLICANT: Kay, B. K.
:
: APPLICANT: Fowkes, D. M.
:
: TITLE OF INVENTION: Totally Synthetic Affinity Reagents
:
: NUMBER OF SEQUENCES: 186
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Pennie & Edmonds
:
: STREET: 1155 Avenue of the Americas
:
: CITY: New York
:
: STATE: New York
:
: COUNTRY: U.S.A.
:
: ZIP: 10036-2711
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC Compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471.927
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,331
FILING DATE: 31-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-471-927-106

Query Match 53.4%; Score 31; DB 8; Length 38;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPOG 10
: | | | |
DB 24 NPFTYHPOG 32

RESULT 6
US-08-660-451A-12
Sequence 12, Application US/08660451A
GENERAL INFORMATION:
APPLICANT: Elliott, Kathryn J.
APPLICANT: Harpold, Michael M.
TITLE OF INVENTION: HUMAN NEURONAL NICOTINIC ACETYLCHOLINE
RECEPTOR COMPOSITIONS AND METHODS EMPLOYING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Hallier & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,451A
FILING DATE: June 7, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/484,722
FILING DATE: 06/07/95
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-9370B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 446 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
US-08-660-451A-12

Query Match 53.4%; Score 31; DB 10; Length 446;
Best Local Similarity 46.7%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ADPXXYXPOGXXXVG 15
: | | | | | | | |
DB 163 ADISGYIPNGEDVVG 177

RESULT 7
US-09-248-796-18170
Sequence 18170, Application US/09248796
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796
CURRENT FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 28206
SEQ ID NO 18170
LENGTH: 338
TYPE: PRT
ORGANISM: Candida albicans
FEATURE:
NAME/KEY: UNSURE
LOCATION: (2)
OTHER INFORMATION: Identity of amino acid sequences at the above locations are un
US-09-248-796-18170

Query Match 53.4%; Score 31; DB 16; Length 338;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 DPXXYXPOG 10
: | | | | |
DB 203 DPSQYGPNG 211

RESULT 8
US-09-270-767-41428
Sequence 41428, Application US/09270767
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 41428
LENGTH: 137
TYPE: PRT
ORGANISM: Drosophila melanogaster
FEATURE:
OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-41428


```
Query Match          53.4%; Score 31; DB 16; Length 137;
Best Local Similarity 60.0%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADPXXYXPOG 10
   |||
Db 87 ADENGYQPG 96

RESULT 9
US-09-328-352-6263
; Sequence 6263, Application US/09328352
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6263
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6263

Query Match          53.4%; Score 31; DB 17; Length 261;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PXXYXPOG 10
   |||
Db 80 PAAYHPOG 87

RESULT 10
US-09-096-409-18170
; Sequence 18170, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/09/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 18170
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Candida albicans
; NAME/KEY: UNSURE
; LOCATION: (2)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-096-409-18170

Query Match          53.4%; Score 31; DB 19; Length 338;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 DPXXYXPOG 10
   |||
Db 203 DPSQYGPNG 211

RESULT 11
US-60-138-830-326
; Sequence 326, Application US/60138830
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED ION CHANNEL PROTEINS, NUCLEIC
```

```
; TITLE OF INVENTION: ACID MOLECULES ENCODING ION CHANNEL PROTEINS, AND USES THERE
; FILE REFERENCE: CL000030
; CURRENT APPLICATION NUMBER: US/60/138,830
; CURRENT FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 400
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 326
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Drosophila
US-60-138-830-326

Query Match          53.4%; Score 31; DB 19; Length 121;
Best Local Similarity 38.5%; Pred. No. 50;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 PXXYXPOGXXXVG 15
   |||
Db 84 PKDIPTGATVIG 96

RESULT 12
US-60-140-804-711
; Sequence 711, Application US/60140804
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED ION CHANNEL PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ION CHANNEL PROTEINS, AND USES
; FILE REFERENCE: CL000040
; CURRENT APPLICATION NUMBER: US/60/140,804
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 1028
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 711
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Drosophila
US-60-140-804-711

Query Match          53.4%; Score 31; DB 19; Length 249;
Best Local Similarity 38.5%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 PXXYXPOGXXXVG 15
   |||
Db 146 PKDIPTGATVIG 158

RESULT 13
US-60-140-956-2074
; Sequence 2074, Application US/60140956
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED SECRETED PROTEINS, NUCLEIC ACID
; FILE REFERENCE: CL000041
; CURRENT APPLICATION NUMBER: US/60/140,956
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 2638
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2074
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Drosophila
US-60-140-956-2074

Query Match          53.4%; Score 31; DB 19; Length 60;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

QY 1 ADPXXYPQG 10
|||
Db 46 ADENGYQPG 55

RESULT 14
US-60-140-956-2309
; Sequence 2309, Application US/60140956
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED SECRETED PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES ENCODING SECRETED PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO00041
; CURRENT APPLICATION NUMBER: US/60/140.956
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 2638
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2309
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Drosophila
US-60-140-956-2309

Query Match 53.4%; Score 31; DB 19; Length 79;
Best Local Similarity 60.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADPXXYPQG 10
|||
Db 46 ADENGYQPG 55

RESULT 15
US-60-143-753-271
; Sequence 271, Application US/60143753
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED ION CHANNEL PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ION CHANNEL PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00046
; CURRENT APPLICATION NUMBER: US/60/143.753
; CURRENT FILING DATE: 1999-07-01
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(393)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-60-143-753-271

Query Match 53.4%; Score 31; DB 19; Length 393;
Best Local Similarity 38.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 PXXYPQGXVVG 15
|||
Db 290 PKDIPTGTVIG 302

Search completed: November 13, 1999, 05:08:39
Job time: 10865 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:08:04 ; Search time 75.45 Seconds
(without alignments)
7.965 Million cell updates/sec

Title: US-08-913-430-11
Perfect score: 58
Sequence: 1 ADPXXYPQGXVVG 15

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	33	56.9	329	2	S25246	mob protein - Bord
2	32	55.2	1019	2	A32856	collagen alpha 1(V
3	31	53.4	1670	1	CGHU3B	collagen alpha 3(I
4	31	53.4	295	2	C69180	adhesion protein -
5	31	53.4	123	2	T01778	hypothetical prote
6	31	53.4	119	2	S78093	endocuticular prot
7	30	51.7	1396	1	VCBE40	major capsid prote
8	30	51.7	885	2	S76357	penicillin-binding
9	30	51.7	868	2	D69297	transmembrane olig
10	30	51.7	925	2	T00781	hypothetical prote
11	30	51.7	192	2	S56309	probable membrane
12	30	51.7	1603	2	S17983	gene posterior sex
13	30	51.7	644	2	A53184	myc far upstream e
14	30	51.7	1458	2	A45665	adult-specific bru
15	29.5	50.9	792	2	T00082	hypothetical prote
16	29	50.0	1376	1	VCBED6	major capsid prote
17	29	50.0	621	2	JC5164	acetolactate synth
18	29	50.0	454	2	A64468	glutamate-ammonia
19	29	50.0	372	2	S23936	adhesion molecule
20	29	50.0	106	2	A27227	amyloid protein A
21	29	50.0	122	2	B3843	serum amyloid prot
22	29	50.0	122	2	A23843	serum amyloid prot
23	29	50.0	122	2	I49496	amyloid A - mouse
24	29	50.0	83	2	I71951	serum amyloid A -
25	29	50.0	660	2	A44432	amino acid transpo
26	29	50.0	485	2	D65118	pantothenate perme
27	29	50.0	444	2	S35783	glycoprotein gx -
28	29	50.0	627	2	T00124	hypothetical prote
29	29	50.0	331	2	B65106	probable proteinas
30	29	50.0	715	2	H71489	probable transcrip
31	29	50.0	211	2	S75866	hypothetical prote
32	29	50.0	132	2	T02116	hypothetical prote
33	29	50.0	635	2	T01695	disulfite reductase
34	29	50.0	748	2	S66129	angiotensinogen
35	29	50.0	256	2	H71157	hypothetical prote
36	28	48.3	455	1	OOCG	rhodopsin - giant
37	28	48.3	440	1	S60755	rhodopsin - Allote
38	28	48.3	452	1	S34332	rhodopsin - northe
39	28	48.3	448	1	S29483	rhodopsin - Japane

ALIGNMENTS

RESULT 1

S25246

mob protein - Bordetella bronchiseptica plasmid pBBR1

C:Species: Bordetella bronchiseptica

C>Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 17-Mar-1999

C:Accession: S25246

R:Antoine, R.; Locht, C.

Mol. Microbiol. 6, 1785-1799, 1992

A:Title: Isolation and molecular characterization of a novel broad-host-range plasmid

A:Reference number: S25245; MUID:92334155

A:Accession: S25246

A:Molecule type: DNA

A:Residues: 1-329 <ANT>

A:Cross-references: EMBL:X66730

A:Note: the authors translated the codon GAC for residue 267 as His

C:Genetics:

A:Gene: mob

A:Genome: plasmid

Query Match 56.9%; Score 33; DB 2; Length 329;
Best Local Similarity 55.6%; Pred. No. 12;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYPQG 10

! | | | |

Db 204 EPRAYAPQG 212

RESULT 2

A32856

collagen alpha 1(VI) chain precursor - chicken

C:Species: Gallus gallus (chicken)

C>Date: 22-Nov-1989 #sequence_revision 22-Nov-1989 #text_change 08-Sep-1997

C:Accession: A32856; I50587; I50627; I50586

R:Bonaldo, P.; Russo, V.; Eucclotti, F.; Bressan, G.M.; Colombatti, A.

J. Biol. Chem. 264, 5573-5580, 1989

A:Title: Alpha-1 chain of chick type VI collagen. The complete cDNA sequence reveals

A:Reference number: A32856; MUID:89174602

A:Accession: A32856

A:Molecule type: mRNA

A:Residues: 1-1019 <BON>

A:Cross-references: GB:J04598; NID:9576463; PID:g211354

A:Note: 479-Asn and 620-Asn was also found

R:Walchli, C.; Koller, E.; Trueb, J.; Trueb, B.

Eur. J. Biochem. 205, 583-589, 1992

A:Title: Structural comparison of the chicken genes for alpha 1(VI) and alpha 2(VI) c

A:Reference number: I50587; MUID:92241293

A:Accession: I50587

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1019 <WAL>

A:Cross-references: EMBL:X57998; NID:g62874; PID:g62875

A:Accession: I50627

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1019 <WA2>

A:Cross-references: EMBL:X64458; NID:g63301; PID:g63302

R:Koller, E.; Trueb, B.

Eur. J. Biochem. 208, 769-774, 1992

A:Title: Characterization of the chicken alpha 1(VI) collagen promoter.

A:Reference number: I50586; MUID:93011107

A:Accession: I50586
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-75 <COL>
A:Cross-references: EMBL:X57987; NID:962872; PID:962873
C:Genetics:
A:Gene: Col6A1
A:Introns: 33/1; 76/2; 143/2; 196/3; 237/3; 244/3; 251/3; 266/3; 284/3; 299/3; 308/3; 3106/1; 650/3; 687/2; 748/3; 807/1; 817/1
C:Superfamily: collagen alpha 1(VI) chain; von Willebrand factor type A repeat homology
C:Keywords: cell binding; coiled coil; disulfide bond; extracellular matrix; glycoprotein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1019/Product: collagen alpha 1(VI) chain #status predicted <MAT>
F:35-213/Domain: von Willebrand factor type A repeat homology <VWA1>
F:255-590/Domain: collagenous #status predicted <COL>
F:440-442/Region: cell attachment (R-G-D) motif
F:476-478/Region: cell attachment (R-G-D) motif
F:529-531/Region: cell attachment (R-G-D) motif
F:611-788/Domain: von Willebrand factor type A repeat homology <VWA2>
F:822-989/Domain: von Willebrand factor type A repeat homology <VWA3>
F:212,514,535,666,799,887/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.2%; Score 32; DB 2; Length 1019;
Best Local Similarity 55.6%; Pred. NO. 63;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXXPQG 10
||| | | | |
DB 361 DPGAYGPKG 369

RESULT 3
CGHU3B
collagen alpha 3(IV) chain precursor, long splice form - human
N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 26-Feb-1999
C:Accession: A54763; A43928; A44043; A45971; A39786
R:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Tryggvason, K.; Reenders, S.T.
J. Biol. Chem. 269, 23013-23017, 1994
A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
A:Reference number: A54763; MUID:94364994
A:Accession: A54763
A:Molecule type: mRNA
A:Residues: 1-1670 <MAR>
A:Cross-references: GB:X80031; NID:9577563; PID:9577564
A:Experimental source: Kidney
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the alpha3(IV) chain.
A:Reference number: A43928; MUID:92147878
A:Accession: A43928
A:Molecule type: mRNA
A:Residues: 1331-1524, 1', 1526-1670 <TUR>
A:Cross-references: GB:M81379
A:Experimental source: Kidney
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 267, 15780-15784, 1992
A:Title: Exon/Intron structure of the human alpha 3(IV) gene encompassing the Goodpasture antigen.
A:Reference number: A44043; MUID:93015826
A:Accession: A44043
A:Molecule type: DNA
A:Residues: 1386-1670 <QUI>
A:Cross-references: GB:M92993; NID:9177895; PID:9177896
A:Note: sequence extracted from NCBI backbone (NCBIP:115597)
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 269, 17358, 1994
A:Reference number: A44738
A:Contents: annotation: erratum; correction to intronic sequence in A44043
R:Bernal, D.; Quinones, S.; Saus, J.
J. Biol. Chem. 268, 12090-12094, 1993

A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
A:Reference number: A45971; MUID:93280184
A:Accession: A45971
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1427-1444 <BER>
R:Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Reenders, S.T.
Am. J. Hum. Genet. 49, 545-554, 1991
A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain
A:Reference number: A39786; MUID:91353570
A:Accession: A39786
A:Molecule type: mRNA
A:Residues: 1453-1593, 'A', 1595-1670 <MOR>
A:Cross-references: GB:555790; NID:9234418; PID:9234419
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit and subsequently O-glycosylated.
C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope of the Goodpasture antigen.
C:Genetics:
A:Gene: GDB:COL4A3
A:Cross-references: GDB:128351; OMIM:120070
A:Map position: 2q36-q37
A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands of the same chromosome.
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3(IV) and one alpha 4(IV) chains. The alpha 3(IV) chain is thought to form a heterotrimer of two alpha 3(IV) and one alpha 4(IV) chains. The alpha 3(IV) chain is thought to form a heterotrimer of two alpha 3(IV) and one alpha 4(IV) chains.
F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MA>
F:29-42/Domain: amino-terminal nonhelical, NHI <NHI>
F:43-1438/Region: interrupted helical
F:791-793/Region: cell attachment (R-G-D) motif
F:996-998/Region: cell attachment (R-G-D) motif
F:1154-1156/Region: cell attachment (R-G-D) motif
F:1306-1308/Region: cell attachment (R-G-D) motif
F:1345-1347/Region: cell attachment (R-G-D) motif
F:1432-1434/Region: cell attachment (R-G-D) motif
F:1439-1670/Domain: carboxyl-terminal nonhelical, NCL <NCL>
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CTR>
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>
F:31, 33, 39, 41, 125, 422, 476, 479, 682, 722, 809, 1387/Disulfide bonds: interchain #status predicted
F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1460-1548, 1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F:1505-1511, 1616-1622/Disulfide bonds: #status predicted
F:1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 53.4%; Score 31; DB 1; Length 1670;
Best Local Similarity 60.0%; Pred. NO. 1.7e+02;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADPXXXPQG 10
| | | | |
DB 613 AGPPGYPQG 622

RESULT 4
C69180
adhesion protein - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
C:Accession: C69180
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicalre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwanli, K.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: C69180

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-295 <MTH>
A:Cross-references: GB:AE000842; GB:AE000666; NID:g2621676; PID:g2621684
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH604
C:Superfamily: adhesin B

Query Match 53.4%; Score 31; DB 2; Length 295;
Best Local Similarity 55.6%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ADPXXYXPQ 9
||| | |
Db 67 ADPHTYPE 75

RESULT 5
T01778
hypothetical protein A_IG002P16.14 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 19-Feb-1999
C:Accession: T01778
R:Miller, N.; Beck, C.; Kramer, J.
Submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana IG002P16.
A:Reference number: 214421
A:Accession: T01778

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-123 <MIL>
A:Cross-references: EMBL:AF007270; NID:g2191157; PID:g2191165
C:Genetics:
A:Map position: 5
A:Introns: 58/2
A:Note: A_IG002P16.14

Query Match 53.4%; Score 31; DB 2; Length 123;
Best Local Similarity 62.5%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PXXYXPQ 10
| | | | |
Db 51 PSAYLPQ 58

RESULT 6
S78093
endocuticular protein SgAbd-3 - desert locust
C:Species: Schistocerca gregaria (desert locust)
C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 12-Dec-1997
C:Accession: S78093
R:Andersen, S.
Submitted to the Protein Sequence Database, November 1997
A:Description: Amino acid sequence studies on endocuticular proteins from the desert locust
A:Reference number: S78091
A:Accession: S78093
A:Molecule type: protein
A:Residues: 1-119 <AND>
A:Experimental source: strain albino; adult; abdominal cuticle
C:Keywords: blocked amino end; glycoprotein; pyroglutamic acid
F:1-119/Product: endocuticular protein SgAbd-3 #status experimental <MAT>
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:96/Binding site: N-acetylglactosamine (Thr) #status experimental

Query Match 53.4%; Score 31; DB 2; Length 119;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ADPXXYXPQ 10
|| | | |
Db 82 ADENGYVPQ 91

RESULT 7
VCBE40
major capsid protein - human herpesvirus 3
C:Species: human herpesvirus 3, varicella-zoster virus
C:Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 14-Nov-1997
C:Accession: E27341
R:Davison, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A:Title: The complete DNA sequence of varicella-zoster virus.
A:Reference number: A27345; MUID:86306657
A:Accession: E27341
A:Molecule type: DNA
A:Residues: 1-1396 <DAV>
A:Cross-references: EMBL:X04370; NID:g59989; PID:g60029
C:Genetics:
A:Gene: 40
C:Superfamily: varicella-zoster virus major capsid protein
C:Keywords: capsid protein

Query Match 51.7%; Score 30; DB 1; Length 1396;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 DPXXYXPQ 10
|| | | |
Db 448 DPQFPQ 456

RESULT 8
S76357
penicillin-binding protein 1A - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein sll0002
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
C:Accession: S76357
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76357
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-885 <KAN>
A:Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PID:d1010860; PID:g100158
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: pona
A:Start codon: GTG

Query Match 51.7%; Score 30; DB 2; Length 885;
Best Local Similarity 35.7%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ADPXXYXPQXXXV 14
|| | | | :
Db 218 ADVVTYAPRGMTI 231

RESULT 9
D69297
transmembrane oligosaccharyl transferase homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998

C:Accession: D69297
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343
A:Accession: D69297
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-868 <KLE>
A:Cross-references: GB:AE001078; GB:AE000782; NID:g3689401; PID:g2650253; TIGR:AF0380

Query Match 51.7%; Score 30; DB 2; Length 868;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 DPXXXPQG 10
|||
Db 69 DPFTYIPYG 77
|||

RESULT 10
T00781
hypothetical protein T22J18.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
C:Accession: T00781
R:Vystotskaia, V.S.; Schwartz, J.R.; Toriumi, M.; Yu, G.; Oji, O.; Kwan, A.; Liu, S.; Li,
riz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, K
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC T22J18 sequence, complete sequence.
A:Reference number: Z14202
A:Accession: T00781
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-925 <VYS>
A:Cross-references: EMBL:AC003979; NID:g3172156; PID:g3287695
C:Genetics:
A:Map position: 1
A:Introns: 44/3; 113/1; 168/2; 242/3; 268/2; 331/1; 363/3; 416/2; 461/2; 594/3; 635/2; 6
A:Note: T22J18.20

Query Match 51.7%; Score 30; DB 2; Length 925;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXXPQ 9
|||
Db 76 DPVVYSPQ 83
|||

RESULT 11
S56309
probable membrane protein YFR054c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein F014
C:Species: Saccharomyces cerevisiae
C:Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 05-Dec-1997
C:Accession: S56309; S62265; S63801
R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanu
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces ce
A:Reference number: S56186
A:Accession: S56309
A:Molecule type: DNA
A:Residues: 1-192 <MUR>
A:Cross-references: EMBL:D50617; NID:g936685; PID:d1009934; PID:g936809; MIPS:YFR054c
R:Murakami, Y.
submitted to the EMBL Data Library, December 1994

A:Reference number: S62230
A:Accession: S62265
A:Molecule type: DNA
A:Residues: 1-192 <MUW>
A:Cross-references: EMBL:D44597; NID:g871938; PID:d1008610; PID:g871953
R:Eki, T.; Naitou, M.; Hagiwara, H.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, M.; Tsuchiy
Yeast 12, 149-167, 1996
A:Title: Analysis of a 36.2 kb DNA sequence including the right telomere of chromosome
A:Reference number: S63787; MUID:96287652
A:Accession: S63801
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-192 <EKI>
A:Cross-references: EMBL:D44597; NID:g871938; PID:d1008610; PID:g871953
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995
C:Genetics:
A:Map position: 6R
A:Note: YFR054c
C:Keywords: Transmembrane protein
F:70-86/Domain: transmembrane #status predicted <TM>

Query Match 51.7%; Score 30; DB 2; Length 192;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 5; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADPXXXPQG 10
|:| | | |
Db 117 AEPGYPVG 126
|:| | | |

RESULT 12
SI7983
gene posterior sex combs protein - fruit fly (Drosophila melanogaster)
N:Alternate names: regulatory protein Psc
C:Species: Drosophila melanogaster
C:Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 05-Dec-1998
C:Accession: SI7983; SI7880
R:Brunk, B.P.; Martin, E.C.; Adler, P.N.
Nature 353, 351-353, 1991
A:Title: Drosophila genes Posterior Sex Combs and Suppressor two of zeste encode prot
A:Reference number: SI7983; MUID:92018190
A:Accession: SI7983
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1603 <BRU>
A:Cross-references: EMBL:X59275; NID:g8357; PID:g8358
A:Note: translation of nucleotide sequence is not complete
R:van Lohuizen, M.; Frasch, M.; Wientjens, E.; Berns, A.
Nature 353, 353-355, 1991
A:Title: Sequence similarity between the mammalian bmi-1 proto-oncogene and the Dros
A:Reference number: SI7880; MUID:92018191
A:Accession: SI7880
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 231-452, 'I', 454-591 <LOH>
C:Genetics:
A:Gene: FlyBase:Psc
A:Cross-references: FlyBase:FBgn0005624
C:Superfamily: RING finger homology
C:Keywords: DNA binding; nucleus
F:261-309/Domain: RING finger homology <RNG>

Query Match 51.7%; Score 30; DB 2; Length 1603;
Best Local Similarity 38.5%; Pred. No. 2.6e+02;
Matches 5; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 PXXXPQGXVVG 15
| | | | | | | | | |
Db 912 PHLYGPKGTWNG 924
| | | | | | | | | |

RESULT 13
A53184
myc far upstream element-binding protein - human
N:Alternate names: FUSE-binding protein
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 10-Sep-1997
C:Accession: A53184
R:Duncan, R.; Bazar, L.; Michelotti, G.; Tomonaga, T.; Kruttsch, H.; Avigan, M.; Levens, G.; Dev. 8, 465-480, 1994
A:Title: A sequence-specific, single-strand binding protein activates the far upstream element
A:Reference number: A53184; MUID:94170991
A:Accession: A53184
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-644 <DUN>
A:Cross-references: GB:U05040; NID:9460151; PID:9460152
C:Keywords: DNA binding

Query Match 51.7%; Score 30; DB 2; Length 644;
Best Local Similarity 62.5%; Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 PXXYXPQG 10
| | | | |
Db 502 PAPIAPQG 509
RESULT 14
A45665
adult-specific brush border esterase/phospholipase (EC 3.-.-.-) precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 10-Sep-1997
C:Accession: A45665
R:Boll, W.; Schmid-Chanda, T.; Semenza, G.; Mantel, N.
J. Biol. Chem. 268, 12901-12911, 1993
A:Title: Messenger RNAs expressed in intestine of adult but not baby rabbits. Isolation and characterization of the complementary DNAs

A:Reference number: A45665
A:Accession: A45665
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1458 <BOL>
A:Cross-references: GB:Z12841; NID:91689; PID:91690
C:Keywords: hydrolase

Query Match 51.7%; Score 30; DB 2; Length 1458;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 DPXXYXPQ 9
| | | | |
Db 521 DPVRYSPQ 528
RESULT 15
T00082
hypothetical protein KIAA0517 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 22-Jan-1999
C:Accession: T00082
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The completed sequence of the cDNA for human KIAA0517
A:Reference number: Z14086
A:Accession: T00082
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-792 <NAG>
A:Cross-references: EMBL:AB011089; NID:dl185359; PID:dl026373
A:Experimental source: brain; clone HG0752
C:Genetics:

A>Note: KIAA0517
Query Match 50.9%; Score 29.5; DB 2; Length 792;
Best Local Similarity 70.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 ADPXXYXPQG 10
| | | | |
Db 758 ADP-LYGPQG 766
Search completed: November 13, 1999, 12:08:06
Job time: 2078 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:33:41 ; Search time 51.07 Seconds
(without alignments)
8.303 Million cell updates/sec

Title: US-08-913-430-11

Perfect score: 58

Sequence: 1 ADPXXYPGXXVVG 15

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	32	55.2	1019	1	CA16_CHICK
2	31	53.4	712	1	ARN2_MOUSE
3	31	53.4	1670	1	CA34_HUMAN
4	31	53.4	143	1	CUL7_BOMMO
5	31	53.4	127	1	SAA_MACEU
6	30	51.7	1458	1	PHLX_RABIT
7	30	51.7	1603	1	PSC_DRONE
8	30	51.7	423	1	YFET_PSEAE
9	30	51.7	1396	1	VCAP_VZVD
10	30	51.7	364	1	YFET_PSEAE
11	30	51.7	192	1	YFET_PSEAE
12	29	50.0	467	1	CBPA_DICDI
13	29	50.0	454	1	GLNA_METJA
14	29	50.0	621	1	ILVB_MYCAV
15	29	50.0	372	1	LEMI_RAT
16	29	50.0	956	1	NUT1_MAGGR
17	29	50.0	483	1	PANF_ECOLI
18	29	50.0	122	1	SAAL_MOUSE
19	29	50.0	122	1	SAAL_MOUSE
20	29	50.0	102	1	SAA5_MESAU
21	29	50.0	127	1	SAA5_ANAPL
22	29	50.0	660	1	SL54_PIG
23	29	50.0	1376	1	VCAP_HSVB
24	29	50.0	444	1	VGLX_HSVBS
25	29	50.0	331	1	YHBU_ECOLI
26	28	48.3	763	1	AM02_HUMAN
27	28	48.3	269	1	BAS1_HUMAN
28	28	48.3	743	1	BGAL_THEET
29	28	48.3	295	1	CFXQ_CYAME
30	28	48.3	503	1	CP33_HUMAN
31	28	48.3	502	1	CP34_HUMAN
32	28	48.3	503	1	CP35_HUMAN
33	28	48.3	503	1	CP37_HUMAN
34	28	48.3	503	1	CP38_MACFA
35	28	48.3	503	1	CP3L_CALJA
36	28	48.3	547	1	CP78_MAIZE
37	28	48.3	210	1	CUC30_BOMMO
38	28	48.3	123	1	CYC2_CAEEL
39	28	48.3	762	1	E13E_TRIHA
40	28	48.3	1308	1	ERB4_HUMAN
41	28	48.3	222	1	ERB4_RAT
42	28	48.3	900	1	FOX2_YEAST
43	28	48.3	384	1	GLF1_KLEPN

ALIGNMENTS

RESULT	ID	Score	Query Match %	Length	ID	Description
1	CA16_CHICK	58	55.2	1019	1	CA16_CHICK
2	AC	57.85	53.4	712	1	ARN2_MOUSE
3	DT	57.85	53.4	1670	1	CA34_HUMAN
4	DT	57.85	53.4	143	1	CUL7_BOMMO
5	DT	57.85	53.4	127	1	SAA_MACEU
6	DE	57.85	53.4	1458	1	PHLX_RABIT
7	GN	57.85	53.4	1603	1	PSC_DRONE
8	OS	57.85	53.4	423	1	YFET_PSEAE
9	OC	57.85	53.4	1396	1	VCAP_VZVD
10	OC	57.85	53.4	364	1	YFET_PSEAE
11	OC	57.85	53.4	192	1	YFET_PSEAE
12	RN	57.85	53.4	467	1	CBPA_DICDI
13	RP	57.85	53.4	454	1	GLNA_METJA
14	RP	57.85	53.4	621	1	ILVB_MYCAV
15	RP	57.85	53.4	372	1	LEMI_RAT
16	RA	57.85	53.4	956	1	NUT1_MAGGR
17	RL	57.85	53.4	483	1	PANF_ECOLI
18	RL	57.85	53.4	122	1	SAAL_MOUSE
19	RL	57.85	53.4	122	1	SAAL_MOUSE
20	RA	57.85	53.4	102	1	SAA5_MESAU
21	RA	57.85	53.4	127	1	SAA5_ANAPL
22	RL	57.85	53.4	660	1	SL54_PIG
23	CC	57.85	53.4	1376	1	VCAP_HSVB
24	CC	57.85	53.4	444	1	VGLX_HSVBS
25	CC	57.85	53.4	331	1	YHBU_ECOLI
26	CC	57.85	53.4	763	1	AM02_HUMAN
27	CC	57.85	53.4	269	1	BAS1_HUMAN
28	CC	57.85	53.4	743	1	BGAL_THEET
29	CC	57.85	53.4	295	1	CFXQ_CYAME
30	CC	57.85	53.4	503	1	CP33_HUMAN
31	CC	57.85	53.4	502	1	CP34_HUMAN
32	CC	57.85	53.4	503	1	CP35_HUMAN
33	CC	57.85	53.4	503	1	CP37_HUMAN
34	CC	57.85	53.4	503	1	CP38_MACFA
35	CC	57.85	53.4	503	1	CP3L_CALJA
36	CC	57.85	53.4	547	1	CP78_MAIZE
37	CC	57.85	53.4	210	1	CUC30_BOMMO
38	CC	57.85	53.4	123	1	CYC2_CAEEL
39	CC	57.85	53.4	762	1	E13E_TRIHA
40	CC	57.85	53.4	1308	1	ERB4_HUMAN
41	CC	57.85	53.4	222	1	ERB4_RAT
42	CC	57.85	53.4	900	1	FOX2_YEAST
43	CC	57.85	53.4	384	1	GLF1_KLEPN

Query Match 55.2%; Score 32; DB 1; Length 1019;
Best Local Similarity 55.6%; Pred. No. 46;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 2 DPXXYPQG 10
   |||
Db 361 DPGAYGPKG 369

RESULT 2
ARN2_MOUSE
ID ARN2_MOUSE STANDARD; PRT; 712 AA..
AC Q61324;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR 2 (ARNT PROTEIN 2).
GN ARNT2.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MORIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE: 96239527.
RA HIROSE K., MORITA M., EMA M., MIMURA J., HAMADA H., FUJII H.,
RA SAJO Y., GOTOH O., SOGAWA K., FUJII-KURIYAMA Y.;
RT "CDNA cloning and tissue-specific expression of a novel basic
RT helix-loop-helix/PAS factor (Arnt2) with close sequence similarity
RT to the aryl hydrocarbon receptor nuclear translocator (Arnt).";
RL MOL. CELL. BIOL. 16:1706-1713(1996).
CC -1- FUNCTION: SPECIFICALLY RECOGNIZES THE XENOBIOTIC RESPONSE ELEMENT
CC (ARE).
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH AN OTHER
CC BHLH PROTEIN. HETERODIMER WITH THE ARYL HYDROCARBON RECEPTOR (AHR)
CC OR THE SIM1 PROTEIN.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- TISSUE SPECIFICITY: RESTRICTED TO ADULT BRAIN AND KIDNEY.
CC -1- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. HIGHEST TO THE ARNT PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63644; GI304146; --
DR MGD; MGI:107188; ARNT2.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
DR PFAM; PF00010; HLH; 1.
DR PFAM; PF00989; PAS; 1.
KW REPEAT; DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION.
FT DNA_BIND 64 76
FT DOMAIN 137 117 BASIC DOMAIN.
FT REPEAT 137 204 PAS-1.
FT REPEAT 325 391 PAS-2.
FT DOMAIN 398 441 PAC MOTIF.
FT DOMAIN 73 76 POLY-ARG.
FT DOMAIN 440 445 POLY-GLN.
FT DOMAIN 552 555 POLY-SER.
SQ SEQUENCE 712 AA; 77884 MW; 47C6A1B7 CRC32;
```

Query Match 53.4%; Score 31; DB 1; Length 712;
Best Local Similarity 62.5%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 ADPXXYP 8
   |||
Db 603 ADPSSVSP 610

RESULT 3
CA34_HUMAN
ID CA34_HUMAN STANDARD; PRT; 1670 AA.
AC Q01955;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PROCOLLAGEN ALPHA 3(IV) CHAIN PRECURSOR.
GN COL4A3.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE: 94364994.
RA MARIYAMA M., LEINONEN A., MOCHIZUKI T., TRYGGVASON K., REEDERS S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain.
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT human tissues.";
RL J. BIOL. CHEM. 269:23013-23017(1994).
RN [2]
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE: 93013826.
RA QUINONES S., BERNAL D., GARCIA-SOGO M., ELENA S.F., SAUS J.;
RT "Exon/Intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT antigenic region at the triple helix/NC1 domain junction.";
RL J. BIOL. CHEM. 267:19780-19784(1992).
RN [3]
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE: 91353570.
RA MORRISON K.E., MARIYAMA M., YANG-FENG T.L., REEDERS S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL AM. J. HUM. GENET. 49:545-554(1991).
RN [4]
RP SEQUENCE OF 1331-1670 FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE: 92147878.
RA TURNER N., MASON P.J., BROWN R., FOX M., POVEY S., REES A.,
RA PUSEY C.D.;
RT "Molecular cloning of the human Goodpasture antigen demonstrates it
RT to be the alpha 3 chain of type IV collagen.";
RL J. CLIN. INVEST. 89:592-601(1992).
RN [5]
RP SEQUENCE OF 1644-1670 FROM N.A.
RC TISSUE=KIDNEY;
RA DING J.;
RA SUBMITTED (JAN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [6]
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC TISSUE=KIDNEY;
RX MEDLINE: 94124597.
RA FENG L., XIA Y., WILSON C.B.;
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)
RT collagen gene. Differential expression of mRNA transcripts that
RT predict three protein variants with distinct carboxyl regions.";
RL J. BIOL. CHEM. 269:2342-2348(1994).
RN [7]
RP ALTERNATIVE SPLICING.
RX MEDLINE: 93280184.
RA BERNAL D., QUINONES S., SAUS J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced.";
RL J. BIOL. CHEM. 268:12090-12094(1993).
RN [8]
RP VARIANT PRO-1474.
```



```
CC -----
DR EMBL: AB004766; D1021308;
DR PROSITE: PS00233; CUTICLE; 1.
DR PFAM: PF00379; insect cuticle; 1.
KW STRUCTURAL PROTEIN; CUTICLE; SIGNAL.
FT SIGNAL 1 16
FT CHAIN 17 143
SQ SEQUENCE 143 AA; 15270 MW; 73807C10 CRC32;

Query Match 53.4%; Score 31; DB 1; Length 143;
Best Local Similarity 60.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADPXXYXPG 10
DB 94 ADENGYPQG 103

RESULT 5
ID SAA_MACEU STANDARD; PRT; 127 AA.
AC P53613.
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DE SERUM AMYLOID A PROTEIN PRECURSOR (SAA).
OS MACROPUS EUGENII (TAMMAR WALLABY).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC DIPROTODONTIA; MACROPODIDAE; MACROPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=LIVER;
RX MEDLINE: 96179383.
RA UHLAR C.M., BLACK I.L., SHIELDS D.C., BRACK C.M., SCHREIBER G.,
RA WHITEHEAD A.S.;
RT "wallaby serum amyloid A protein: cDNA cloning, sequence and
RT evolutionary analysis.";
RL SCAND. J. IMMUNOL. 43:271-276(1996).
CC -!- FUNCTION: MAJOR ACUTE PHASE REACTANT. APOLIPOPROTEIN OF THE HDL
CC COMPLEX.
CC -!- TISSUE SPECIFICITY: PLASMA, SYNTHESIZED BY THE LIVER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39363; G1136126;
DR PROSITE: PS00992; SAA; 1.
DR PFAM: PF00277; SAA_proteins; 1.
KW ACUTE PHASE; PLASMA; HDL; AMYLOID; SIGNAL.
FT SIGNAL 1 18
FT CHAIN 19 127
SQ SEQUENCE 127 AA; 14322 MW; 511396FF CRC32;

Query Match 53.4%; Score 31; DB 1; Length 127;
Best Local Similarity 55.6%; Pred. No. 9.6;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPG 10
DB 114 DPNRYRPG 122

RESULT 6
PHLX_RABIT
```

```
ID PHLX_RABIT STANDARD; PRT; 1458 AA.
AC Q05017;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PHOSPHOLIPASE ADRA-B-B PRECURSOR (EC 3.1.-.-).
OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX TISSUE=INTESTINE;
RX MEDLINE: 93286138.
RA BOLL W., SCHMID-CHANDA T., SEMENZA G., MANTEI N.;
RT "Messenger RNAs expressed in intestine of adult but not baby rabbits.
RT Isolation of cognate cDNAs and characterization of a novel brush
RT border protein with esterase and phospholipase activity.";
RL J. BIOL. CHEM. 268:12901-12911(1993).
CC -!- FUNCTION: HAS ESTERASE AND PHOSPHOLIPASE A/LYSOPHOSPHOLIPASE
CC ACTIVITY. CAN CONVERT PHOSPHATIDYLCHOLINE TO FATTY ACIDS AND
CC GLYCEROPHOSPHOCHOLINE. COULD BE INVOLVED IN UPTAKE OF DIETARY
CC LIPIDS, POSSIBLY INCLUDING LONG CHAIN RETINYL ESTERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. BRUSH BORDER.
CC -!- TISSUE SPECIFICITY: INTESTINE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE INTESTINE OF ADULT BUT NOT
CC BABY RABBITS.
CC -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z12841; G1690;
DR PIR: A45665; A45665.
DR PROSITE: PS01098; LIPASE_GDSL_SER; 2.
DR PFAM: PF00657; Lipase_GDSL; 3.
KW HYDROLASE; REPEAT; SIGNAL; TRANSMEMBRANE.
FT SIGNAL 1 19
FT CHAIN 20 1458
FT DOMAIN 20 1415
FT TRANSMEM 1416 1439
FT DOMAIN 1440 1458
FT DOMAIN 39 1403
FT REPEAT 39 347
FT REPEAT 362 707
FT REPEAT 708 1054
FT REPEAT 1064 1403
FT ACT_SITE 400 400
FT ACT_SITE 747 747
FT ACT_SITE 1103 1103
SQ SEQUENCE 1458 AA; 161343 MW; B155E7F3 CRC32;

Query Match 51.7%; Score 30; DB 1; Length 1458;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPG 9
DB 521 DPNRYSPQ 528

RESULT 7
PSC_DROME
ID PSC_DROME STANDARD; PRT; 1603 AA.
AC P35820;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
```

DE POSTERIOR SEX COMBS PROTEIN.
GN PSC.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 92018190.
RA BRUNK B.P., MARTIN E.C., SHARP E., ADLER P.N.;
RT "Drosophila genes posterior Sex Combs and Suppressor two of zeste
encode proteins with homology to the murine bmi-1 oncogene.";
RL NATURE 353:351-353(1991).
CC
CC EXPRESSION PATTERNS OF THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN
CC ANTENNAPEDIA (ANTP-C) AND BITHORAX (BX-C) COMPLEXES, AND HENCE FOR
CC THE MAINTENANCE OF SEGMENTAL DETERMINATION.
CC
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59275; G8358; .
DR PIR; S17983; S17983.
DR FLYBASE; FBgn0005624; PSC.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR PFAM; PF00097; zf-C3HC4; 1.
KW ZINC-FINGER; DEVELOPMENTAL PROTEIN; DNA-BINDING; NUCLEAR PROTEIN.
FT DOMAIN 83 88
FT DOMAIN 91 98 POLY-THR.
FT DOMAIN 145 150 POLY-THR.
FT DOMAIN 184 202 POLY-SER.
FT DOMAIN 642 651 POLY-SER.
FT ZN_FING 265 303 C3HC4-TYPE.
FT SEQUENCE 1603 AA; 169999 MW; B5B35A43 CRC32;
SQ

Query Match 51.7%; Score 30; DB 1; Length 1603;
Best Local Similarity 38.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 PXXYXPGQXXXVG 15
| | | | | : |
Db 912 PHLGPRGETKMG 924

RESULT 8
PYR2_PSEAE STANDARD; PRT; 423 AA.
AC Q51551;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DIHYDROOROTASE-LIKE PROTEIN (ASPARTATE CARBAMOYLTRANSFERASE 44 KD NON-
DE CATALYTIC CHAIN).
GN PYR' OR PYRX.
OS PSEUDOMONAS AERUGINOSA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RA VICKREY J.F., SCHURR M.J., BENJAMIN R.C., CUNIN R., SHANLEY M.S.,
RA O'DONOVAN G.A.;
RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: NON-FUNCTIONAL DHOASE.
CC
CC -1- SUBUNIT: HETERODODECAMER OF 6 ACTIVE PYR SUBUNITS AND 6 NON-

CC CATALYTIC PYR' SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: TO DIHYDROOROTASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L19649; G309886; .
DR PFAM; PF00744; Dihydroorotase; 1.
KW PYRIMIDINE BIOSYNTHESIS.
FT DOMAIN 81 84 POLY-ALA.
SQ SEQUENCE 423 AA; 44136 MW; 7BC6CCD9 CRC32;

Query Match 51.7%; Score 30; DB 1; Length 423;
Best Local Similarity 40.0%; Pred. No. 50;
Matches 6; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 ADPXXYXPGQXXXVG 15
| | | | | : |
Db 375 ADLVLFDPQGSTLAG 389

RESULT 9
VCAP_VZVD STANDARD; PRT; 1396 AA.
ID VCAP_VZVD
AC P09245;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE MAJOR CAPSID PROTEIN (MCP).
GN 40.
OS VARICELLA-ZOSTER VIRUS (STRAIN DUMAS) (VZV).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 86306657.
RA DAVISON A.J., SCOTT J.E.;
RT "The complete DNA sequence of varicella-zoster virus.";
RL J. GEN. VIROL. 67:1759-1816(1986).
CC -1- FUNCTION: MAJOR PROTEIN OF THE ICOSAEDRAL CAPSID.
CC
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES MAJOR CAPSID PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X04370; G60029; .
DR PIR; E27341; VCBE40.
KW COAT PROTEIN.
SQ SEQUENCE 1396 AA; 154979 MW; 10F752F9 CRC32;

Query Match 51.7%; Score 30; DB 1; Length 1396;
Best Local Similarity 55.6%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPG 10
| | | : |
Db 448 DPRQPPPG 456

RESULT 10
YFIT_PSEAE

Query Match 50.0%; Score 29; DB 1; Length 467;
 Best Local Similarity 62.5%; Pred. No. 89;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PXXYXPOG 10
 DB 245 PGAYPQG 252

RESULT 13
 ID GLNA_METJA STANDARD; PRT; 454 AA.
 AC Q60182;
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE GLUTAMINE SYNTHETASE (EC 6.3.1.2) (GLUTAMATE-AMMONIA LIGASE) (GS).
 GN GLNA OR MJ1346
 OS METHANOCOCCUS JANNASCHII.
 OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
 CC METHANOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE: 96337999.
 RA BULT C.-J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
 RA SCOTT J.L., GEORGEHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii";
 RL SCIENCE 273:1058-1073(1996).
 CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) -> ADP + GLUTAMINE +
 CC ORTHOPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U67574; G1591988; -.
 DR TIGR: MJ1346; -.
 DR PROSITE: PS00180; GLNA_1; 1.
 DR PROSITE: PS00181; GLNA_ATP; 1.
 DR PFAM: PF00120; gln-synt; 1.
 DR HSP: P06201; 2LGS.
 KW LIGASE.
 SQ SEQUENCE 454 AA; 51398 MW; AEACE20D CRC32;

Query Match 50.0%; Score 29; DB 1; Length 454;
 Best Local Similarity 44.4%; Pred. No. 87;
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPOG 10
 DB 265 EFSFYDEG 273

RESULT 14
 ID ILVB_MYCAV STANDARD; PRT; 621 AA.
 AC Q59498;
 DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE PROBABLE ACETOLACTATE SYNTHASE (EC 4.1.3.18) (ACETOHYDROXY-ACID
 DE SYNTHASE) (ALS).
 GN ILVB.
 OS MYCOBACTERIUM AVIUM.
 CC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
 CC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA GUSBERTI L., CANTONI R., DE ROSSI E., BRANZONI M., RICCARDI G.;
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
 CC -!- CATALYTIC ACTIVITY: 2-ACETOLACTATE + CO(2) -> 2-PYRUVATE.
 CC -!- COFACTOR: THIAMINE PYROPHOSPHATE, AND MAGNESIUM ION (BY
 CC SIMILARITY).
 CC -!- PATHWAY: FIRST STEP IN VALINE AND ISOLEUCINE BIOSYNTHESIS.
 CC -!- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L49392; G1196507; -.
 DR PROSITE: PS00187; TPP_ENZYMES; 1.
 DR PFAM: PF00205; TPP_ENZYMES; 1.
 KW BRANCHED-CHAIN AMINO ACID BIOSYNTHESIS; FLAVOPROTEIN; MAGNESIUM;
 KW THIAMINE PYROPHOSPHATE; LYASE.
 FT ACT_SITE 88 BY SIMILARITY.
 SQ SEQUENCE 621 AA; 65913 MW; E11BE73E CRC32;

Query Match 50.0%; Score 29; DB 1; Length 621;
 Best Local Similarity 38.5%; Pred. No. 1.2e+02;
 Matches 5; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 PXXYXPOGXXXVG 15
 DB 396 PLSYGPQSDGSLG 408

RESULT 15
 ID LEM1_RAT STANDARD; PRT; 372 AA.
 AC P30836;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE L-SELECTIN PRECURSOR (LYMPH NODE HOMING RECEPTOR) (LEUKOCYTE ADHESION
 DE MOLECULE-1) (LAM-1) (LY-22) (LYMPHOCYTE SURFACE MEL-14 ANTIGEN)
 DE (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 1) (LECAM1) (CD62L).
 GN SELL OR LNH OR LY-22.
 OS RATTUS NORVEGICUS (RAT).
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92329548.
 RA WATANABE T., SONG Y., HIRAYAMA Y., TAMATANI T., KUIDA K., MIYASAKA M.;
 RT "Sequence and expression of a rat cdna for LECAM-1";
 RL BIOCHIM. BIOPHYS. ACTA 1131:321-324(1992).
 CC -!- FUNCTION: CELL SURFACE ADHESION PROTEIN. MEDIATE THE ADHERENCE
 CC OF LYMPHOCYTES TO ENDOTHELIAL CELLS OF HIGH ENDOTHELIAL
 CC VENULES IN PERIPHERAL LYMPH NODES.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- SIMILARITY: TO OTHER SELECTINS/LECAMs.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 SUSHI (SCR) REPEATS.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D10831; G220802; -
DR PIR; S23936; S23936
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
DR PFAM; PF00008; EGF; 1.
DR PFAM; PF00059; lectin_c; 1.
DR PFAM; PF00084; sushi; 2.
DR HSSP; P14151; IKJB.
DR CELL ADHESION; TRANSMEMBRANE; GLYCOPROTEIN; EGF-LIKE DOMAIN; LECTIN;
KW SELECTIN; SIGNAL; SUSHI; REPEAT.
FT SIGNAL 1 28 BY SIMILARITY.
FT PROPEP 29 38 BY SIMILARITY.
FT CHAIN 39 372 L-SELECTIN.
FT DOMAIN 39 332 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 333 355 POTENTIAL.
FT DOMAIN 356 372 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 55 155 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 156 192 EGF-LIKE.
FT DOMAIN 196 317 2 X SUSHI (SCR) REPEATS.
FT REPEAT 196 255 SUSHI 1.
FT REPEAT 258 317 SUSHI 2.
FT DISULFID 57 155 BY SIMILARITY.
FT DISULFID 128 147 BY SIMILARITY.
FT DISULFID 160 171 BY SIMILARITY.
FT DISULFID 165 180 BY SIMILARITY.
FT DISULFID 182 191 BY SIMILARITY.
FT DISULFID 197 241 BY SIMILARITY.
FT DISULFID 227 254 BY SIMILARITY.
FT DISULFID 259 303 BY SIMILARITY.
FT DISULFID 289 316 BY SIMILARITY.
FT CARBOHYD 60 60 POTENTIAL.
FT CARBOHYD 104 104 POTENTIAL.
FT CARBOHYD 177 177 POTENTIAL.
FT CARBOHYD 226 226 POTENTIAL.
FT CARBOHYD 246 246 POTENTIAL.
FT CARBOHYD 278 278 POTENTIAL.
SQ SEQUENCE 372 AA; 42441 MW; 6492CFC4 CRC32;

Query Match 50.0%; Score 29; DB 1; Length 372;
Best Local Similarity 62.5%; Pred. No. 71;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXXYXPQ 9
|| | |
DB 183 DPXYGPQ 190

Search completed: November 13, 1999, 10:33:42
Job time: 5192 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 18:59:59 ; Search time 104.22 Seconds
(without alignments)
4.091 Million cell updates/sec

Title: US-08-913-430-12

Perfect score: 87
Sequence: 1 AGXLQNSLLEEVWYLAL 18

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	85	97.7	18	1 W01024	Mycoplasma 72-75 k
2	64	73.6	627	1 W62451	Mycoplasma hyopneu
3	40	46.0	553	1 W03528	Salmonella enteric
4	39.5	45.4	979	1 W40309	Human ITAK protein
5	39	44.8	282	1 P90422	Human phenylethano
6	39	44.8	1455	1 W48663	Fanconi anemia of
7	37	42.5	2408	1 R24306	Translation of ORF
8	37	42.5	3712	1 R13896	ACV synthetase. DN
9	37	42.5	3639	1 R40227	ACVS. DNA coding d
10	36	41.4	745	1 R76110	Human A.D. New nuc
11	36	41.4	745	1 W03560	Human adrenoleucod
12	36	41.4	89	1 W27781	Amino acid sequenc
13	36	41.4	391	1 W62299	Synechocystis D1
14	35	40.2	515	1 P91235	(ENV-80)(GAG-VII)
15	35	40.2	291	1 P80297	Sequence encoded b
16	35	40.2	96	1 P70239	Polypeptide ENV(80
17	35	40.2	3054	1 R40841	Translation of TEV
18	35	40.2	972	1 R92288	Cycloisomaltoligo
19	35	40.2	934	1 R92289	Sarcolectin fragme
20	35	40.2	135	1 W23821	Human sarcolectin.
21	35	40.2	469	1 W23820	Human sarcolectin.
22	35	40.2	973	1 W41815	A. thaliana trehal
23	35	40.2	473	1 W56451	UDP-glucose:flavon
24	35	40.2	47	1 W61457	NRCIVIE construct
25	34	39.1	582	1 W22499	Phafla derived ca
26	34	39.1	875	1 W33660	Human puromycin-se
27	34	39.1	920	1 W33662	Murine puromycin-s
28	34	39.1	825	1 W33661	Human puromycin-se
29	34	39.1	602	1 W33663	Human puromycin-se
30	34	39.1	905	1 W36879	Human DNA repair e
31	34	39.1	1233	1 W34536	Ndaurelia beta vi
32	34	39.1	1233	1 W41935	Ndaurelia beta-11
33	34	39.1	878	1 W54373	IBDV RNA-dependent
34	34	39.1	686	1 W60040	Cotton cellulose s
35	34	39.1	1039	1 W73309	Cellulose synthase
36	34	39.1	64	1 W85466	Secreted protein e
37	34	39.1	878	1 W95397	IBDV VPI protein.
38	34	39.1	878	1 W95393	IBDV VPI protein.
39	34	39.1	423	1 W93359	WO 99/07855 SeqID
40	34	39.1	419	1 W93360	Human regulatory m
41	34	39.1	188	1 W93946	Hemidesmosome-prom
42	33.5	38.5	285	1 R63580	Rat matrix protein
43	33.5	38.5	295	1 R94374	

44 33.5 38.5 583 1 W30704 Mouse hyaluronate
45 33.5 38.5 543 1 W36503 Human hyaluronate

ALIGNMENTS

```
RESULT 1
W01024
ID W01024 standard; Peptide; 18 AA.
AC W01024;
DT 19-JAN-1997 (first entry)
DE Mycoplasma 72-75 kda protective antigen N-terminal peptide.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumonia;
KW diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
FH Key Location/Qualifiers
FT misc_difference 3
FT /note= "undetermined amino acid"
PN W09628472-Al.
PD 19-SEP-1996.
PR 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (UYME ) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
DR WPI; 96-433763/43.
PT Putative protective antigens against Mycoplasma - used for the
PT detection, prevention or treatment of Mycoplasma infections, esp. M.
PT hyopneumoniae in swine
PS Claim 7; Page 27; 43pp; English.
CC A 72-75 kda putative protective antigen against Mycoplasma
CC contains the N-terminal sequence given in W01024 and the internal
CC CNBr fragments given in W01025-27. The antigen was isolated from
CC Mycoplasma hyopneumoniae cells using antibody probes enriched with
CC Mycoplasma-specific antibodies. Other protective antigens were
CC also identified (see also W01028-37). Protective antigens and
CC antibodies can be used in vaccines for preventing or treating
CC mycoplasma infections, partic. M. hyopneumoniae infections in
CC swine. They can also be used for diagnosis.
SQ Sequence 18 AA;

Query Match 97.7%; Score 85; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGXLQNSLLEEVWYLAL 18
   |||||
DB 1 AGXLQNSLLEEVWYLAL 18

RESULT 2
W62451
ID W62451 standard; Protein; 627 AA.
AC W62451;
DT 01-OCT-1998 (first entry)
DE Mycoplasma hyopneumoniae P65 surface antigen.
KW Mycoplasma hyopneumoniae; surface lipoprotein P65; structural gene;
KW antigen; vaccine; swine; enzootic pneumoniae; diagnosis; pig;
KW immune response; mycoplasma pneumoniae.
OS Mycoplasma hyopneumoniae.
PN US788962-A.
PD 04-AUG-1998.
PR 28-AUG-1996; 703947.
PR 17-JAN-1995; US-373957.
PR 28-AUG-1996; US-703947.
PA (UMOR ) UNIV MISSOURI.
DR WPI; 98-446004/38.
DR N-PSDB; V39862.
PT Vaccine against mycoplasma pneumoniae in pigs - containing
PT Mycoplasma hyopneumoniae P65 surface antigen fusion protein
PT Claim 1; Fig 2; 29pp; English.
```

CC A vaccine has been developed for protecting pigs against mycoplasma
 CC pneumonia caused by Mycoplasma hyopneumoniae. The vaccine contains an
 CC immunogenic fusion protein comprising a first amino acid sequence fused
 CC to a second amino acid sequence, where the first sequence is the present
 CC sequence of 627 amino acids, which is a P65 surface antigen. The
 CC sequence given in V39862 represents the Mycoplasma hyopneumoniae
 CC surface lipoprotein P65 structural gene, which encodes the P65 surface
 CC antigen. The vaccine is used to induce an immune response in pigs
 CC against mycoplasma pneumonia caused by Mycoplasma hyopneumoniae.
 SQ Sequence 627 AA;

Query Match 73.6%; Score 64; DB 1; Length 627;
 Best Local Similarity 83.3%; Pred. No. 0.0033;
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGXLOKNSLLEEVWYAL 18
 ||||| |||||
 Db 28 AGCLOKNSLSEVNYAL 45

RESULT 3
 W03528
 ID W03528 standard; Protein; 553 AA.
 AC W03528;
 DT 25-FEB-1997 (first entry)
 DE Salmonella enterica IagA protein.
 KW IagA; IagB; Salmonella enterica; primer; probe; HeLa.
 OS Synthetic.
 PN EP-721989-A1.
 PD 17-JUL-1996.
 PF 15-JAN-1996; 400098.
 PR 16-JAN-1995; FR-000410.
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PI (INSP) INST PASTEUR.
 PI Le GUERN FELLOUS M, Popoff MY;
 DR WPI: 96-322837/33.
 DR N-PSDB; T37466.
 DT New nucleic acid of S. enterica spp. enterica involved in cell
 PT invasion - and derived oligo-nucleotide(s) useful as primers and
 PT probes for detecting Salmonella in food etc.
 PS Claim 13; Fig 1; 33pp; French.

CC This is the amino acid sequence of the IagA protein from Salmonella
 CC enterica spp. enterica serovar typhi. The protein has a calculated
 CC mol. wt. of 63026 Da. The N-terminal portion of the protein has homology
 CC with a similar region of the transcriptional regulatory proteins Phob
 CC (24% identity and 52% similarity) and Phop (25% identity and 62%
 CC similarity) from E.coli. The IagA and IagB sequences are used to
 CC generate primers and probes (T37451-65) which are useful for detecting
 CC S. enterica and/or S. bongori in samples e.g. in food. The IagA and IagB
 CC proteins are involved in invasion of cultured HeLa cells by S. enterica.
 SQ Sequence 553 AA;

Query Match 46.0%; Score 40; DB 1; Length 553;
 Best Local Similarity 54.5%; Pred. No. 40;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 LOKNSLLEEVW 14
 :|||:||||
 Db 55 VSKNILLDQWV 65

RESULT 4
 W40309
 ID W40309 standard; Protein; 979 AA.
 AC W40309;
 DT 23-JUN-1998 (first entry)
 DE Human ITAK protein.
 KW Interleukin-1/tumour necrosis factor alpha activated kinase: ITAK; human;
 KW inhibitor; cytokine mediated inflammation; antagonist; disorder; therapy;
 KW IL-1; TNF.
 OS Homo sapiens.

PN W09747750-A1.
 PD 18-DEC-1997.
 PF 09-JUN-1997; U08516.
 PR 10-JUN-1996; US-633414.
 PA (IMV) IMMUNEX CORP.
 PI Anderson DM, Bird TA, Sims JE, Virca G;
 DR N-PSDB; V10514; V10525.
 DR Nucleic acid encoding interleukin-1/tumour necrosis factor activated
 PT kinase - used to identify specific antagonists for treatment of
 PT cytokine-mediated inflammation
 PS Claim 3; Fig 1; 80pp; English.
 CC This sequence represents an interleukin-1/tumour necrosis factor alpha
 CC activated kinase (ITAK). Antagonists of ITAK are used to treat IL-1 or
 CC TNF alpha-mediated inflammatory disorders e.g. rheumatoid arthritis,
 CC inflammatory bowel disease, type I diabetes, psoriasis, Alzheimer's
 CC disease, reperfusion injury, malignancy, transplant rejection, neuropathy
 CC associated with human immunodeficiency virus etc. Cells containing ITAK
 CC can be used to raise antibodies for assay of ITAK or to inhibit IL-1 and
 CC TNF alpha activity. Gene products that associate with ITAK are potential
 CC inhibitors and can be used to detect ITAK genes. Antisense sequences
 CC inhibit expression of ITAK. Inhibition of ITAK selectively blocks cell
 CC responses to IL-1 and TNF alpha, but not responses to other cytokines.
 SQ Sequence 979 AA;

Query Match 45.4%; Score 39.5; DB 1; Length 979;
 Best Local Similarity 60.0%; Pred. No. 94;
 Matches 9; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

QY 5 QRNSLLEE--VWYL 16
 ||| |||||
 Db 143 QKDKLFEENVWYL 157

RESULT 5
 P90422
 ID P90422 standard; protein; 282 AA.
 AC P90422;
 DT 1-NOV-1989 (first entry)
 DE Human phenylethanolamine N-methyl transferase
 KW Phenylethanolamine N-methyl transferase; human; circulatory
 OS diseases; psychopharmaceuticals.
 PN Homo sapiens (Human)
 PN J01137971-A.
 PD 24-MAY-1989.
 PF 24-NOV-1987; 295458.
 PR 24-NOV-1987; JP-295458.
 PA (FUJI) Fujita Giken KK.
 DR WPI: 89-198217/27.
 DR N-PSDB; N90232.
 PT New phenyl-ethanolamine N-methyl transferase
 PT - used for prepn. of medicines for circulatory system
 PT diseases, etc.
 PS Claim 1; fig 2; 11pp; Japanese.
 CC Phenylethanolamine N-methyl transferase (see N90232), obtd. from
 CC human medulla glandulae suprarenalis-derived pheochromocytoma cells,
 CC human bulb cells, hypothalamus or nervus maxillaris human cells.
 CC used in the prepn. of medicines for circulatory diseases and
 CC psychopharmaceuticals.
 SQ Sequence 282 AA;

Query Match 44.8%; Score 39; DB 1; Length 282;
 Best Local Similarity 56.2%; Pred. No. 28;
 Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 GXLOKNSLLEEVWYLA 17
 ||||| |||||
 Db 209 GHLLIGALEESWYLA 224

RESULT 6

W48663
 ID W48663 standard; Protein: 1455 AA.
 AC W48663;
 DT 28-AUG-1998 (first entry)
 DE Fanconi anaemia of complementation group A protein.
 KW Fanconi anaemia of complementation group A; FA-A; genetic defect;
 KW prenatal FA-A; FA-A carrier detection; disease diagnosis.
 OS Homo sapiens.
 PN W09814462-A1.
 PD 09-APR-1998.
 PF 03-OCT-1997; U18010.
 PR 04-OCT-1996; US-726012.
 PA (FANC-) FANCONI ANEMIA RES FUND INC.
 PI Joenje H, Lo ten Foe JR;
 DR WPI: 98-240012/21.
 DR N-PSDB; V18187.
 PT DNA for Fanconi Anaemia complementation group A - useful for, e.g.
 PT developing products for diagnosis and screening of disease and gene
 PT therapy
 PS Claim 1; Fig 3; 63pp; English.
 CC This sequence is the Fanconi anaemia of complementation group A
 CC (FA-A) protein of the invention. The DNA's may be used to complement a
 CC genetic defect in a cell (especially the FA-A gene). The products can be
 CC used for screening (especially prenatal FA-A), detection of FA-A carriers
 CC and FA-A disease diagnosis.
 SQ Sequence 1455 AA;

Query Match 44.8%; Score 39; DB 1; Length 1455;
 Best Local Similarity 53.8%; Pred. No. 1.8e+02;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 LQKNSLLEEVWYL 16
 : : : : :
 DB 173 IQSSLLLEAWHL 185

RESULT 7
 R24306
 ID R24306 standard; Protein: 2408 AA.
 AC R24306;
 DT 20-NOV-1992 (first entry)
 DE Translation of ORF 2 contg. E.faecium protein VanS.
 KW Glycopeptide antibiotic; vancomycin; teicoplanin; resistant;
 KW D-Ala-D-Ala ligase; peptidoglycan precursor; transposon;
 KW inverted repeats; vanR; vns; vanH; vana; vanX;
 KW open reading frame.
 OS Enterococcus faecium BM417.
 FH Key Location/Qualifiers
 FT protein 727..1115
 FT /label= Vans
 FT /note= "see R24296"
 PN W09207942-A.
 PD 14-MAY-1992.
 PF 29-OCT-1991; F00855.
 PR 31-OCT-1990; FR-013579.
 PA (INSP) INST PASTEUR.
 PI Arthur M, Courvalin P, Dutka-malen S, Molinas C;
 DR WPI: 92-183677/22.
 DR N-PSDB; Q25183.
 PT Polypeptides involved in expression of glycopeptide antibiotic
 PT resistance - useful in diagnosing presence of Gram-positive
 PT enterococcal strains e.g. Enterococcus faecium and E gallinarum
 PS Disclosure; Fig 5; 163pp; French.
 CC A 7.3kb fragment of E.faecium DNA containing the five genes vanH,
 CC vana, vanX, vanR and vns involved in antibiotic resistance was
 CC translated in each of the three possible open reading frames.
 CC Within ORF 2 there is the vns gene. The "X"s in the sequence
 CC indicate the position of nonsense codons.
 CC See also Q25178-Q25182.
 SQ Sequence 2408 AA;

Query Match 42.5%; Score 37; DB 1; Length 2408;
 Best Local Similarity 40.0%; Pred. No. 7.1e+02;
 Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 QY 4 LQKNSLLEEVWYLAL 18
 : : : : :
 DB 2257 LQNRFRKEICYXSI 2271

RESULT 8
 R13896
 ID R13896 standard; Protein: 3712 AA.
 AC R13896;
 DT 22-NOV-1991 (first entry)
 DE ACV synthetase.
 KW Beta lactam antibiotics; penicillin.
 OS Acromonium chrysogenum.
 FH Key Location/Qualifiers
 FT domain 301..1068
 FT /label= I
 FT /function= activation of amino acid substrate
 FT region 374..423
 FT /label= subdomain
 FT region 474..501
 FT /label= subdomain
 FT region 655..699
 FT /label= subdomain
 FT region 725..754
 FT /label= subdomain
 FT domain 1392..2154
 FT /label= II
 FT /function= activation of amino acid substrate
 FT region 1470..1518
 FT /label= subdomain
 FT region 1564..1590
 FT /label= subdomain
 FT region 1745..1789
 FT /label= subdomain
 FT region 1817..1846
 FT /label= subdomain
 FT domain 2474..3295
 FT /label= III
 FT /function= activation of amino acid substrate
 FT region 2554..2603
 FT /label= subdomain
 FT region 2647..2673
 FT /label= subdomain
 FT region 2827..2871
 FT /label= subdomain
 FT region 2899..2928
 FT /label= subdomain
 FT domain 3560..3647
 FT /label= IV
 FT /function= thioesterase
 PN EP-445868-A.
 PD 11-SEP-1991.
 PF 27-FEB-1991; 200423.
 PR 28-FEB-1990; EP-200475.
 PR 28-FEB-1990; EP-200488.
 PR 02-JUL-1990; EP-201768.
 PR 03-OCT-1990; EP-202628.
 PR 27-FEB-1991; EP-200423.
 PA (KONN) GIST-BROCADES NV.
 PI Veenstra AE, Martin JF, Garcia BD, Gutierrez S, Barredo JL;
 PI Montenegro PE, Von Doehren H, Palissa H, Van Liempt H;
 DR WPI: 91-268735/37.
 DR N-PSDB; Q13608.
 PT DNA encoding amino:adipyl-cysteinyI-valine synthetase - used for
 PT prodn. of the enzyme or enhanced prodn. of new or known
 PT beta-lactam antibiotic cpds.
 PS Claim 1; Page 20; 54pp; English.
 CC The DNA sequence was obt'd. from five subclones isolated from a
 CC gene library of A. chrysogenum c10 (ATCC 48). The protein

CC sequence was deduced from the DNA. Three distinct regions of
 CC homology have been identified, domains I, II and III. Within
 CC these domains several even more conserved elements can be
 CC distinguished. Since the enzyme synthesises a tripeptide, which
 CC most probably requires the activation of three amino acids, a
 CC role for these domains in the activation reactions seems likely.
 CC A fourth domain is thought to act as a thioesterase.
 CC The gene can be used to express the synthetase enzyme which can
 CC be used for the prodn. of new beta-lactam antibiotics.
 CC See also R13896.
 CC Sequence 3712 AA;

Query Match 42.5%; Score 37; DB 1; Length 3712;

Best Local Similarity 50.0%; Pred. No. 1-2e+03;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 7 NSLLEEVWYAL 18
 :||: ||| |
 Db 1149 HSIQLTVHVL 1160

RESULT 9

ID R40227 standard; Protein; 3639 AA.
 AC R40227;
 DT 21-FEB-1994 (first entry)
 DE ACVS.
 KW Delta-(L-alpha-aminoacyl)-L-cystinyl-D-valine synthase; ACVS;
 KW beta-lactam; antibiotic; transformed; cephalosporin; vector.
 OS Acrononium chrysogenum.
 PN J05192162-A.
 PD 03-AUG-1993.
 PF 25-JUL-1991; 186222.
 PR 31-JUL-1990; JP-205677.
 PA (TAKE) TAKEDA CHEM IND LTD.
 DR WPI; 93-277475/35.
 DR N-PSDB; Q48231.
 PT DNA coding delta-(L-alpha-aminoacyl)-L-cystinyl D-valine
 PT synthase - for improved productivity of cephalosporin antibiotics
 PS Claim 1; Page 14-27; 69pp; Japanese.
 CC The sequence (Q48231) is of a vector which includes the
 CC delta-(L-alpha-aminoacyl)-L-cystinyl-D-valine synthase gene.
 CC This sequence was transformed into a host cell to express the ACVS
 CC product. The protein produced (R40227) was then used to manufacture
 CC a beta-lactam antibiotic.
 CC Sequence 3639 AA;

Query Match 42.5%; Score 37; DB 1; Length 3639;

Best Local Similarity 50.0%; Pred. No. 1.1e+03;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 7 NSLLEEVWYAL 18
 :||: ||| |
 Db 1076 HSIQLTVHVL 1087

RESULT 10

ID R76110 standard; Protein; 745 AA.
 AC R76110;
 DT 16-NOV-1995 (first entry)
 DE Human ALD.
 KW ALD gene; adrenoleukodystrophy; adrenomyeloneuropathy; AMN;
 KW diagnosis; gene therapy.
 OS Homo sapiens.
 PN CA2108606-A.
 PD 16-APR-1995.
 PF 15-OCT-1993; 108606.
 PR 15-OCT-1993; CA-108606.
 PA (INRM) INST NAT SANTE & RECH MEDICALE.
 PI Aubourg P, Mandel J, Mosser J, Sarde C;

DR WPI; 95-215721/29.
 DR N-PSDB; Q94048.
 PT New nucleic acid responsible for adreno-leuco-dystrophy - related
 PT probes, proteins and antibodies, useful for diagnosis and treatment
 PS Claim 16; Fig.2A-B; 39pp; English.
 CC Probes corresp. to breakpoints in the red pigment gene of an AMN
 CC patient were used to isolate clones from an Xq28 cosmid library.
 CC Following hybridizations and nested PCR, a HeLa cell cDNA library
 CC was screened, and a complete ALD cDNA fragment (Q94048) encoding the
 CC protein given in R94048 was obt'd. Transformation of hematopoietic
 CC cells with ALD-encoding sequences in vivo will allow therapy of ALD
 CC or AMN.
 CC Sequence 745 AA;

Query Match 41.4%; Score 36; DB 1; Length 745;

Best Local Similarity 60.0%; Pred. No. 2.8e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 9 LLEEVWYAL 18
 ||| :||: |
 Db 321 LLERLWYML 330

RESULT 11

ID W03560 standard; Protein; 745 AA.
 AC W03560;
 DT 26-FEB-1997 (first entry)
 DE Human adrenoleucodystrophy wild-type protein ALDP.
 KW Adrenoleucodystrophy; gene therapy; retroviral vector M48;
 KW adrenomyelopathy; membrane protein; long chain fatty acid oxidation.
 OS Homo sapiens.
 PN W09621733-A2.
 PD 18-JUL-1996.
 PF 12-JAN-1996; F00059.
 PR 13-JAN-1995; FR-000376.
 PA (INRM) INST NAT SANTE & RECH MEDICALE.
 PI Aubourg P, Mandel JL, Mosser J, Sarde CO;
 DR WPI; 96-342286/34.
 DR N-PSDB; T39335.
 PT Recombinant viral vector contg. DNA for correcting
 PT adrenoleuco-dystrophy - and immunologically tolerable cells contg.
 PT this vector, useful in gene therapy
 PS Example; Fig 7: 36pp; French.
 CC A 2.43 kb SpeI-EcoRI fragment of human wild-type ALD
 CC (adrenoleucodystrophy) cDNA was inserted into retroviral vector M48
 CC so that it was under control of the mouse phosphoglycerate kinase
 CC (PGK) promoter. The resulting vector was co-transfected with vector
 CC PGK-neomycin into amphotropic packaging cell line psiCRIP. Neomycin
 CC (G418)-resistant transfectant cells which were also positive with
 CC anti-ALD antibodies were incubated with skin fibroblasts of an ALD
 CC patient. After two rounds of infection, about 70% of cells were
 CC expressing normal ALD protein (ALDP) and integration of M48-ALD was
 CC confirmed by Southern blotting. The present sequence is that of
 CC ALDP.

CC Sequence 745 AA;

Query Match 41.4%; Score 36; DB 1; Length 745;

Best Local Similarity 60.0%; Pred. No. 2.8e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 9 LLEEVWYAL 18
 ||| :||: |
 Db 321 LLERLWYML 330

RESULT 12

ID W27781 standard; Protein; 89 AA.
 AC W27781;
 DT 21-JUL-1998 (first entry)

DE Amino acid sequence of an outer membrane protein F precursor.
 KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
 KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 OS Staphylococcus aureus.
 FH Key Location/Qualifiers
 FT Misc_difference 1. .89 /note= "residues designated X are not defined in
 FT the specification"
 FT Misc_difference 50 /note= "encoded by GGN"
 FT W09730070-AL.
 PN 21-AUG-1997.
 PD 19-FEB-1997; U02318.
 PR 20-FEB-1996; US-011888.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Black Mt, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
 DR WPI: 97-424969/39.
 DR N-PSDB: T83750.
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against S.
 PT aureus infection
 PS Claim 6; Page 271; 989pp; English.
 CC The present sequence represents a Staphylococcus aureus protein, that,
 CC based on homology with an Escherichia coli protein, is believed to be an
 CC outer membrane protein F precursor. The DNA sequence was isolated from a
 CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
 CC sequence can be used in the construction of ribozymes and antisense
 CC sequences to control the expression of Staphylococcal genes. The DNA
 CC sequence is also useful as a source of regulatory elements for the
 CC control of bacterial gene expression. The present protein may be used
 CC to produce vaccines to enable a host to produce specific antibodies
 CC with antibacterial action. These vaccines and antibodies would protect
 CC a host against infection by S. aureus, and conditions relating to
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.
 SQ Sequence 89 AA;

Query Match 41.4%; Score 36; DB 1; Length 89;
 Best Local Similarity 35.3%; Pred. NO. 25;
 Matches 6; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GXLOKNSLLEEVWYVAL 18
 | | | | | : : : : :
 Db 69 GETQXNSALDKLMSVGI 85

RESULT 13
 ID W62299 standard; Protein; 391 AA.
 AC W62299;
 DT 28-SEP-1998 (first entry)
 DE Synecocystis D1 protease.
 KW Scenedesmus; D1 protease; herbicide; inhibition; alga; wheat;
 KW detection; identification.
 OS Synecocystis Sp.
 FH Key Location/Qualifiers
 FT Misc_difference 8 /note= "encoded by CTC"
 FT Misc_difference 20 /note= "encoded by GAC"
 FT Misc_difference 21 /note= "encoded by CAC"
 FT Misc_difference 173 /note= "encoded by AAA"
 FT Misc_difference 197 /note= "encoded by TCC"
 FT Misc_difference 213 /note= "encoded by AAA"
 FT Misc_difference 232 /note= "encoded by AAA"
 FT Misc_difference 232

Query Match 41.4%; Score 36; DB 1; Length 391;
 Claim 23; Page 50-51; 73pp; English.
 CC A method has been developed for identifying a herbicidal agent which
 CC inhibits D1 protease. The method comprises: (a) incubating a D1 protease
 CC enzyme in a sample suspected of containing a herbicidal agent with a
 CC suitable D1 enzyme substrate, where an enzyme product is formed; and (b)
 CC detecting and quantifying the enzyme product of step (a). Also described
 CC in the present invention are: (i) a method for detecting in vivo a
 CC herbicidal agent which inhibits D1 protease comprising: (a) preparing a
 CC reaction mixture comprising: (i) a wild type cell comprising an active D1
 CC protease enzyme capable of processing a D1 pre-protein; and a Photosystem
 CC II core complex capable of variable chlorophyll fluorescence; and (ii) a
 CC suspected herbicidal agent which inhibits D1 protease; and (iii) suitable
 CC growth medium; (b) illuminating the reaction mixture at illumination
 CC conditions to permit D1 turnover; and (c) measuring the variable
 CC chlorophyll fluorescence produced in step (b), where the level of the
 CC variable chlorophyll fluorescence correlates with herbicidal activity of
 CC the suspected herbicidal agent; (2) a gene encoding a D1 protease enzyme
 CC isolated from: (a) wheat; (b) tobacco; (c) spinach; (d) Scenedesmus; or
 CC (e) Synecocystis; (see V39818 to V39822, which encode W62297 to W62301
 CC respectively). Native or recombinant D1 proteases can be used to develop
 CC assays for the detection of herbicidal compositions capable of
 CC inhibiting D1.
 SQ Sequence 391 AA;

Best Local Similarity 47.1%; Pred. No. 1.3e+02;
 Matches 8; Conservative 3; Mismatches 4; Indels 2; Gaps 1;
 QY 2 CXLQKN--SLLEEVWYL 16
 | | | | | : : : : :
 Db 2 GYLQDNPKELDEVWQI 18

RESULT 14

P91235
 ID P91235 standard; protein; 515 AA.
 AC P91235;
 DT 22-MAR-1990 (first entry)
 DE (ENV-80)(GAG-VII)(Hexahis) protein antigen.
 KW HIV; histidine; antibody; reagent; epitope.
 OS Human immunodeficiency virus.
 PN EP-339389-A.
 PD 02-NOV-1989.
 PF 14-APR-1989; 106688.
 PR 25-APR-1988; CH-001537.
 PR 02-MAY-1988; CH-001635.
 PA (HOFF) F. Hoffmann-La Roche AG.
 PI Dobeil H, Hochuli E, Schacher A;
 DR WPI; 89-31278/44.
 DT Diagnostic reagent for detecting antibodies against viral antigens
 PT - having recombinant viral antigen attached to a solid
 PT phase bonded via metal complexed nitrilo acetic acid deriv.
 PS Disclosure; Fig. 1; 8pp; german.
 CC The (ENV-80)(GAG-VII)(Hexahis) antigen is used in a diagnostic reagent.
 CC It has adjacent His residues and is linked to a glass bead with epoxide
 CC surface functional gps. via a metal ion-loaded nitrilotriacetic acid
 CC deriv. and opt. a spacer. The antigen and its epitopes are bonded in a
 CC well-defined manner, unlike the random epitope arrangement when the
 CC antigens are bonded by adsorption. The reagent is used to detect
 CC antibodies against HIV viral antigens.
 SQ Sequence 515 AA;

Query Match 40.2%; Score 35; DB 1; Length 515;
 Best Local Similarity 62.5%; Pred. No. 2.7e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 7 NSLLEEVW 14
 | | | | : :
 Db 61 NKLEEQIW 68

RESULT 15

P80297
 ID P80297 standard; Protein; 291 AA.
 AC P80297;
 DT 10-MAR-1993 (revised entry)
 DE Sequence encoded by the env (80)-gag-15 gene.
 KW HTLV-III; env gene; gag gene; antigen; vaccine; AIDS antibodies;
 KW Immunocassay; diagnosis; HIV; LAV.
 OS HTLV-III.
 PN EP-270114-A.
 PD 08-JUN-1988.
 PF 03-DEC-1987; 117899.
 PR 05-DEC-1986; GB-029116.
 PA (HOFF) HOFFMANN-LA ROCHE AG.
 PI Gentz R, Le Grice S, Mous J, Stuber D;
 DR WPI; 88-156343/23.
 DR N-PSDB; N80485.
 DT New polypeptide with immunogenic activity of HTLV-3 env-gag
 PT protein - useful in vaccines and for AIDS diagnosis
 PS Claim 1; Page 22 and Fig 9; 49pp; German.
 CC The inventors claim the polypeptide encoded by env (80)-gag-15 gene
 CC (P80297), its fragments and their analogues which differ by AA
 CC substitution(s) and have at least one antigenic and/or immunogenic
 CC determinant of HTLV-III env or gag proteins. The gene encoding
 CC P80297 (N80485) is also claimed. The polypeptides of the invention
 CC are prepd. using vectors prepd. from: plasmid pA-ENV-20, contg. a

CC synthetic env (80) gene, ligated with BamHI-digested pD58/RBSII
 CC (contg. the beta-lactamase gene replication region, a promoter-
 CC operator element, ribosome binding site and dehydrofolate reductase
 CC gene). The product, contg. the env gene in the correct orientation,
 CC is pENV (80)-DHFR. This was cut with BamHI and XhoI and the
 CC resulting fragment ligated with (a) the fragment (contg. the
 CC gag-region of HTLV-III) isolated from pUGAG by sequential treatment
 CC with Hind III, Pvu II and BamHI, and (b) the BamHI-XhoI fragment of
 CC pD55/RBSII, 3A+5A (similar to pD58/RBSII but without the DHFR
 CC gene). The resulting plasmid is pENV (80)-GAG-15. Polypeptide
 CC expression by this plasmid is improved by deleting part of the
 CC cat-gene (from pD58/RBSII) to form pENV (80)-GAG-15 (delta cat).
 SQ Sequence 291 AA;

Query Match 40.2%; Score 35; DB 1; Length 291;
 Best Local Similarity 62.5%; Pred. No. 1.4e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 7 NSLLEEVW 14
 | | | | : :
 Db 61 NKLEEQIW 68

Search completed: November 13, 1999, 19:00:00
 Job time: 139 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:34 ; Search time 64.87 Seconds
(without alignments)
3.171 Million cell updates/sec

Title: US-08-913-430-12

Perfect score: 87

Sequence: 1 AGXLQNSLLEWVYLAL 18

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/PTUS9_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64	73.6	627	US-08-703-947-2	Sequence 2, Appli
2	40	46.0	533	US-08-586-272-2	Sequence 2, Appli
3	39	44.8	1455	US-08-726-012B-2	Sequence 2, Appli
4	37	42.5	3712	US-08-222-617A-4	Sequence 4, Appli
5	37	42.5	3665	US-08-222-617A-13	Sequence 13, Appli
6	37	42.5	3712	US-08-222-617A-25	Sequence 25, Appli
7	36	41.4	745	US-08-136-277-2	Sequence 2, Appli
8	36	41.4	745	US-08-479-403-2	Sequence 2, Appli
9	36	41.4	391	US-08-759-581B-13	Sequence 13, Appli
10	35	40.2	2254	US-08-286-819A-28	Sequence 28, Appli
11	34	39.1	520	US-08-305-505-4	Sequence 4, Appli
12	34	39.1	520	US-08-305-505-5	Sequence 5, Appli
13	34	39.1	520	US-08-305-505-6	Sequence 6, Appli
14	34	39.1	878	US-08-708-541A-26	Sequence 26, Appli
15	34	39.1	188	US-08-933-750C-8	Sequence 8, Appli
16	33.5	38.5	295	US-08-317-223-1	Sequence 1, Appli
17	33.5	38.5	295	PCT-US95-12675-1	Sequence 1, Appli
18	33	37.9	526	US-08-399-986B-4	Sequence 4, Appli
19	33	37.9	526	US-08-493-754A-4	Sequence 4, Appli
20	33	37.9	294	US-08-424-641B-6	Sequence 6, Appli
21	33	37.9	243	US-08-771-455-2	Sequence 2, Appli
22	33	37.9	294	US-08-820-980-6	Sequence 6, Appli
23	32.5	37.4	367	US-08-896-320-1	Sequence 1, Appli
24	32.5	37.4	367	US-08-896-320-3	Sequence 3, Appli
25	32	36.8	184	US-08-033-857A-7	Sequence 7, Appli
26	32	36.8	554	US-08-106-761-2	Sequence 2, Appli
27	32	36.8	422	US-08-036-555B-170	Sequence 170, App
28	32	36.8	422	US-08-469-569-170	Sequence 170, App
29	32	36.8	184	US-08-374-983A-7	Sequence 7, Appli
30	32	36.8	638	US-08-295-814B-13	Sequence 13, Appli
31	32	36.8	770	US-08-445-135-2	Sequence 2, Appli
32	32	36.8	422	US-08-428-926-3	Sequence 3, Appli
33	32	36.8	422	US-08-249-322A-170	Sequence 170, App
34	32	36.8	323	US-08-469-667-16	Sequence 16, Appli
35	32	36.8	473	US-08-597-236-13	Sequence 13, Appli
36	32	36.8	638	US-08-240-783B-2	Sequence 2, Appli
37	32	36.8	312	US-08-240-783B-6	Sequence 6, Appli
38	32	36.8	422	US-08-428-927-3	Sequence 3, Appli
39	32	36.8	422	US-08-428-298-3	Sequence 3, Appli

Sequence 6, Appli
Sequence 3, Appli
Sequence 13, Appli
Sequence 170, App
Sequence 23, Appli
Sequence 2, Appli

40 32 36.8 1528 2 US-08-463-092B-6
41 32 36.8 422 2 US-08-339-517-3
42 32 36.8 473 2 US-08-746-682A-13
43 32 36.8 422 2 US-08-469-526A-170
44 32 36.8 21 2 US-08-508-664-23
45 32 36.8 496 2 US-09-006-675-2

ALIGNMENTS

RESULT 1
US-08-703-947-2
; Sequence 2, Application US/08703947
; Patent No. 5788962
; GENERAL INFORMATION:
; APPLICANT: Wise, Kim S.
; APPLICANT: McIntosh, Mark A.
; TITLE OF INVENTION: DNA Sequences Coding for Mycoplasma
; TITLE OF INVENTION: Hypopneumoniae Surface Antigens,
; TITLE OF INVENTION: Corresponding Proteins and Use in
; TITLE OF INVENTION: Vaccines and Diagnostic Procedures
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace J. Fishel
; STREET: 929 Fee Fee Road, Suite 100
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63043
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
; COMPUTER: Hewlett-Packard Vectra
; OPERATING SYSTEM: MS-DOS version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,947
; FILING DATE: 28-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/373,957
; FILING DATE: January 17, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fishel, Grace J.
; REGISTRATION NUMBER: 25864
; REFERENCE/DOCKET NUMBER: UVM 8141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 878-0440
; TELEFAX: (314) 275-7693
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 amino acid residues
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Predicted amino acid sequence of complete
; DESCRIPTION: 627 residues of the P65 lipoprotein, derived
; DESCRIPTION: from the nucleic acid sequence
; HYPOTHETICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: whole polypeptide
; ORIGINAL SOURCE:
; ORGANISM: Mycoplasma hypopneumoniae
; STRAIN: J
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE: unicellular bacterium
; CELL LINE:
; ORGANELLE:

IMMEDIATE SOURCE:
LIBRARY: Genomic in Charon 4A, GEM12
CLONE: MhpJ25, MhpJ35, MhpJG35, pZJ25, pZJ25.14, pZJ25.14,
CLONE: pZJG35.1, pZJG35.12, pZJG35.13, pZJG35.14
POSITION IN GENOME:
CHROMOSOME/SEGMENT: single chromosome
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: 627 amino acid sequence representing
NAME/KEY: complete sequence (including signal
NAME/KEY: sequence) of surface lipoprotein P65
LOCATION: entire derived coded sequence
IDENTIFICATION METHOD: clone identified by immunodetection of
IDENTIFICATION METHOD: protein product with antiserum specific for
IDENTIFICATION METHOD: P65, residue sequence deduced from nucleic
IDENTIFICATION METHOD: acid sequence
OTHER INFORMATION: immunogenic surface lipoprotein of no known
OTHER INFORMATION: function: C-terminus exposed on external
OTHER INFORMATION: surface of cell; N-terminal signal sequence
OTHER INFORMATION: (first 29 amino acid residues) cleaved during
OTHER INFORMATION: lipid modification process
PUBLICATION INFORMATION:
AUTHORS: Mary F. Kim, Manijeh B. Heidari, Susan J.
AUTHORS: Stull, Mark A. McIntosh, and Kim S. Wise
TITLE: Identification and Mapping of an
TITLE: Immunogenic Region of Mycoplasma
TITLE: hyopneumoniae p65 Surface Lipoprotein
TITLE: Expressed in Escherichia coli from a Cloned
TITLE: Genomic Fragment
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 8
PAGES: 2637-2643
DATE: August 1990
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: From 1 to 627
US-08-703-947-2

Query Match 73.6%; Score 64; DB 2; Length 627;
Best Local Similarity 83.3%; Pred. No. 0.0015;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGXIQKNSLLEEVNYLAL 18
|||
Db 28 AGCLQKNSLLEEVNYLAL 45

RESULT 2
US-08-586-272-2
; Sequence 2, Application US/08586272
; Patent No. 5824795
; GENERAL INFORMATION:
; APPLICANT: POPOFF, MICHEL Y.
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE DETECTION OF
; TITLE OF INVENTION: SALMONELLA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/586,272
FILING DATE: 16-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: FR 9500410
FILING DATE: 16-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 660-107-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-586-272-2

Query Match 46.0%; Score 40; DB 2; Length 553;
Best Local Similarity 54.5%; Pred. No. 20;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 LQKNSLLEEVW 14
:|:|:|:|:|
Db 55 VSKNTLLDQW 65

RESULT 3
US-08-726-012B-2
; Sequence 2, Application US/08726012B
; Patent No. 5952190
; GENERAL INFORMATION:
; APPLICANT: Hans Joenje, et al.
; TITLE OF INVENTION: CDNA FOR FANCONI ANEMIA COMPLEMENTATION GROUP A
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh & Whinston, LLP
; STREET: One World Trade Center, Suite 1600, 121 S.W. Salmon Street
; CITY: Portland
; STATE: OR
; COUNTRY: USA
; ZIP: 97204-2988
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3.5-inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Wordperfect 5.1+, ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,012B
; FILING DATE: 10/04/96
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Richard J. Polley
; REGISTRATION NUMBER: 28,107
; REFERENCE/DOCKET NUMBER: 3812-45520/RJP/DJE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 228-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1455
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: Linear

US-08-726-012B-2

Query Match 44.8%; Score 39; DB 2; Length 1455;
Best Local Similarity 53.8%; Pred. No. 88;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 4 LQKNSLLEEVWYL 16
:|: ||| ||:|
Db 173 IQSSLLEAVVWHL 185

RESULT 4

US-08-222-617A-4
; Sequence 4, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222.617A
; FILING DATE: 04-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 97,157
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3712 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 2555
; OTHER INFORMATION:
; OTHER INFORMATION: /note= "Xaa-Ala or Ser"
US-08-222-617A-4

Query Match 42.5%; Score 37; DB 2; Length 3712;
Best Local Similarity 50.0%; Pred. No. 5.5e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 7 NSLLEEVWYL 18
:|: ||| ||:|
Db 1149 HSIQTWVWHLV 1160

RESULT 5

US-08-222-617A-13

; Sequence 13, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222.617A
; FILING DATE: 04-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 97,157
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3665 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Acremonium chrysogenum
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..3665
; OTHER INFORMATION: /label= ACVS
; OTHER INFORMATION: /note= "ACV Synthetase from Acremonium
; OTHER INFORMATION: chrysogenum; aa 1-3665"
US-08-222-617A-13

Query Match 42.5%; Score 37; DB 2; Length 3665;
Best Local Similarity 50.0%; Pred. No. 5.5e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 7 NSLLEEVWYL 18
:|: ||| ||:|
Db 1149 HSIQTWVWHLV 1160

RESULT 6

US-08-222-617A-25
; Sequence 25, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.

APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222.617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 3712 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-222-617A-25

Query Match 42.5%; Score 37; DB 2; Length 3712;
Best Local Similarity 50.0%; Pred. No. 5.5e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 7 NSLLEVVYLAL 18
:||:|||||
Db 1149 HSIQTVWHLVL 1160

RESULT 7
US-08-136-277-2
Sequence 2, Application US/08136277
Patent No. 5644045
GENERAL INFORMATION:
APPLICANT: MANDEL, Jean-Louis
APPLICANT: AUBOURG, Patrick
APPLICANT: MOSSER, Jean
APPLICANT: SARDE, Claude
TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
TITLE OF INVENTION: CORRESPONDING PROTEIN
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Young & Thompson
STREET: 745 South 23rd Street
CITY: Arlington
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/136.277
FILING DATE: 15-OCT-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: B2272
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 745 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-136-277-2

Query Match 41.4%; Score 36; DB 1; Length 745;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 9 LLEEVVYLAL 18
|||:|||
Db 321 LLERLWVLM 330

RESULT 8
US-08-479-403-2
Sequence 2, Application US/08479403
Patent No. 5869039
GENERAL INFORMATION:
APPLICANT: MANDEL, Jean-Louis
APPLICANT: AUBOURG, Patrick
APPLICANT: MOSSER, Jean
APPLICANT: SARDE, Claude
TITLE OF INVENTION: X-LINKED ADRENOLEUKODYSTROPHY GENE AND
TITLE OF INVENTION: CORRESPONDING PROTEIN
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Young & Thompson
STREET: 745 South 23rd Street
CITY: Arlington
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,403
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: B2272DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 745 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-479-403-2

Query Match 41.4%; Score 36; DB 2; Length 745;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 9 LLEEVWYL 18
||| :||: |
Db 321 LLRLWVWL 330

RESULT 9
US-08-759-581B-13
; Sequence 13, Application US/08759581B
; Patent No. 5876945
; GENERAL INFORMATION:
; APPLICANT: CHISHOLM, DEXTER A.
; APPLICANT: DINER, BRUCE A.
; APPLICANT: DONALDSON, GAIL K.
; APPLICANT: HERSHEY, HOWARD P.
; APPLICANT: JORDAN, DOUGLAS B.
; APPLICANT: TANG, XIAO-SONG
; APPLICANT: TROST, JEFFREY T.
; APPLICANT: WANG, SHAOJIE
; APPLICANT: WARREN, PATRICK V.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING HERBICIDAL AGENTS THAT INHIBIT D1 PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
; SOFTWARE: MICROSOFT WORD 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,581B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-773-0164
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; STRAIN: SYNECHOCYSTIS D1 PROTEINASE PROTEIN
; US-08-759-581B-13

Query Match 41.4%; Score 36; DB 2; Length 391;
Best Local Similarity 47.1%; Pred. No. 68;
Matches 8; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

QY 2 GXLOKN--SLLLEWVYL 16
| | | | | :||: |
Db 2 GYLQDNPKELVDEWQI 18

RESULT 10
US-08-286-819A-28
; Sequence 28, Application US/08286819A
; Patent No. 5871910
; GENERAL INFORMATION:
; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE

; APPLICANT: MOLINAS, CATHERINE
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPOLYPTIDES, IN PARTICULAR
; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,819A
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/174,682
; FILING DATE: 28-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9013579
; FILING DATE: 31-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5871910man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-050-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2254 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-286-819A-28

Query Match 40.2%; Score 35; DB 2; Length 2254;
Best Local Similarity 50.0%; Pred. No. 7, 1e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 LQKNSLLEWVY 15
| | | | | :||: |
Db 2119 LQKNRFEICY 2130

RESULT 11
US-08-305-505-4
; Sequence 4, Application US/08305505
; Patent No. 5668001
; GENERAL INFORMATION:
; APPLICANT: Miziorko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED

;; TITLE OF INVENTION: STABILITY
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Quarles & Brady
;; STREET: 411 East Wisconsin Avenue
;; CITY: Milwaukee
;; STATE: Wisconsin
;; COUNTRY: U.S.A.
;; ZIP: 53202
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/305,505
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/072,040
;; FILING DATE: 02 JUNE 1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Baker, Jean C.
;; REGISTRATION NUMBER: 35,433
;; REFERENCE/DOCKET NUMBER: 65-053-9083-9
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (414) 277-5709
;; TELEFAX: (414) 277-5591
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 520 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-305-505-4

Query Match 39.1%; Score 34; DB 1; Length 520;
Best Local Similarity 60.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 NSLLEEVWYL 16
:| | | | |
Db 446 DSLFEGTWYL 455

RESULT 12
US-08-305-505-5
; Sequence 5, Application US/08305505
; Patent No. 5668001
; GENERAL INFORMATION:
; APPLICANT: Miziorko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/072,040
;; FILING DATE: 02 JUNE 1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Baker, Jean C.
;; REGISTRATION NUMBER: 35,433
;; REFERENCE/DOCKET NUMBER: 65-053-9083-9
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (414) 277-5709
;; TELEFAX: (414) 277-5591
;; INFORMATION FOR SEQ ID NO: 5:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 520 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-305-505-5

Query Match 39.1%; Score 34; DB 1; Length 520;
Best Local Similarity 60.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 NSLLEEVWYL 16
:| | | | |
Db 446 DSLFEGTWYL 455

RESULT 13
US-08-305-505-6
; Sequence 6, Application US/08305505
; Patent No. 5668001
; GENERAL INFORMATION:
; APPLICANT: Miziorko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,505
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,040
; FILING DATE: 02 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 65-053-9083-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 520 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-305-505-6

Query Match 39.1%; Score 34; DB 1; Length 520;
Best Local Similarity 60.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 NSLLEEVWYL 16
:|:|:|:|
Db 446 DSLFEGTWYL 455

RESULT 14

US-08-708-541A-26
; Sequence 26, Application US/08708541A
; Patent No. 5871744
; GENERAL INFORMATION:
; APPLICANT: VAKHARIA, Vikram N.
; APPLICANT: MUNDT, Egbert
; TITLE OF INVENTION: A METHOD FOR GENERATING BIRNAVIRUS FROM
; SYNTHETIC RNA TRANSCRIPTS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAI DO, MARCELSTEIN, MURRAY & ORAM LLP
; STREET: 655 Fifteenth Street, N. W.,
; STREET: Suite 330 - G Street Lobby
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/708,541A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: KITTS, Monica C.
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P8172-6002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/638-5000
; TELEFAX: 202/638-4810
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-708-541A-26

Query Match 39.1%; Score 34; DB 2; Length 878;
Best Local Similarity 50.0%; Pred. No. 3.7e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 SLLEEVWYL 17
:|:|:|:|
Db 305 SMLS DYWILS 314

RESULT 15

US-08-913-750C-8
; Sequence 8, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice

APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 188 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SYNORAB01
CLONE: 77180
US-08-933-750C-8

Query Match 39.1%; Score 34; DB 2; Length 188;
Best Local Similarity 29.4%; Pred. No. 68;
Matches 5; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGXLQKNSLLEEVWYLA 17
:|:|:|:|
Db 63 SGSKENSLLDSIFWMA 79

Search completed: November 13, 1999, 10:56:35
Job time: 1364 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:39 ; Search time 251.81 Seconds
(without alignments)
4.525 Million cell updates/sec

Title: US-08-913-430-12

Perfect score: 87

Sequence: 1 AGXLQNSLLEEVWYAL 18

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA:*

- 1: /cgn2.6/ptodata/1/paa/PTUS9_COMB.pep.*
- 2: /cgn2.6/ptodata/1/paa/US07_COMB.pep.*
- 3: /cgn2.6/ptodata/1/paa/US080_COMB.pep.*
- 4: /cgn2.6/ptodata/1/paa/US081_COMB.pep.*
- 5: /cgn2.6/ptodata/1/paa/US082_COMB.pep.*
- 6: /cgn2.6/ptodata/1/paa/US083_COMB.pep.*
- 7: /cgn2.6/ptodata/1/paa/US084A_COMB.pep.*
- 8: /cgn2.6/ptodata/1/paa/US084B_COMB.pep.*
- 9: /cgn2.6/ptodata/1/paa/US085_COMB.pep.*
- 10: /cgn2.6/ptodata/1/paa/US086_COMB.pep.*
- 11: /cgn2.6/ptodata/1/paa/US087_COMB.pep.*
- 12: /cgn2.6/ptodata/1/paa/US088_COMB.pep.*
- 13: /cgn2.6/ptodata/1/paa/US089_COMB.pep.*
- 14: /cgn2.6/ptodata/1/paa/US090_COMB.pep.*
- 15: /cgn2.6/ptodata/1/paa/US091_COMB.pep.*
- 16: /cgn2.6/ptodata/1/paa/US092_COMB.pep.*
- 17: /cgn2.6/ptodata/1/paa/US093_COMB.pep.*
- 18: /cgn2.6/ptodata/1/paa/US094_COMB.pep.*
- 19: /cgn2.6/ptodata/1/paa/US094_COMB.pep.*
- 20: /cgn2.6/ptodata/1/paa/PTCT_NEW_COMB.pep.*
- 21: /cgn2.6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 22: /cgn2.6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 23: /cgn2.6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 24: /cgn2.6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	97.7	18	13	US-08-913-430-12
2	64	73.6	627	6	US-08-373-957-2
3	43	49.4	400	15	US-09-150-857-6
4	41	47.1	961	1	PCT-US99-16478-11
5	41	47.1	642	17	US-09-390-081-2
6	41	47.1	642	17	US-09-390-081-4
7	40	46.0	140	12	US-08-827-356-5047
8	40	46.0	553	14	US-09-082-969-2
9	39.5	45.4	979	10	US-08-633-414-2
10	39.5	45.4	979	12	US-08-870-529-2
11	39	44.8	483	15	US-09-134-000-4234
12	39	44.8	612	19	US-60-122-952-5
13	38	43.7	147	19	US-60-138-676-545
14	38	43.7	147	19	US-60-140-803-613
15	38	43.7	290	19	US-60-143-993-1840
16	38	43.7	246	23	US-09-417-507-40562
17	37	42.5	3712	2	US-07-658-398A-4
18	37	42.5	3665	2	US-07-658-398A-13
19	37	42.5	3712	2	US-07-658-398B-4

20	37	42.5	3665	2	US-07-658-398B-13
21	37	42.5	705	4	US-08-109-181-4
22	37	42.5	705	4	US-08-142-558-4
23	37	42.5	3712	5	US-08-222-617-4
24	37	42.5	3665	5	US-08-222-617-13
25	37	42.5	705	6	US-08-311-731A-4
26	37	42.5	482	12	US-08-827-356-4910
27	37	42.5	228	12	US-08-827-356-5219
28	37	42.5	549	15	US-09-107-532-7304
29	37	42.5	437	15	US-09-134-000-6150
30	37	42.5	388	15	US-09-134-000-6701
31	37	42.5	92	16	US-09-270-767-34271
32	37	42.5	92	16	US-09-270-767-49488
33	37	42.5	398	19	US-60-146-055-703
34	37	42.5	1171	23	US-09-417-507-41768
35	36	41.4	89	1	PCT-US97-02318-51
36	36	41.4	271	5	US-08-273-257-7
37	36	41.4	391	11	US-08-759-581A-13
38	36	41.4	391	11	US-08-759-581C-13
39	36	41.4	745	12	US-08-835-734-2
40	36	41.4	89	13	US-08-903-470-51
41	36	41.4	944	15	US-09-134-000-5578
42	36	41.4	391	15	US-09-173-281-13
43	36	41.4	97	15	US-09-186-926-35
44	36	41.4	340	16	US-09-248-796-15509
45	36	41.4	632	17	US-09-328-352-4785

ALIGNMENTS

RESULT 1
US-08-913-430-12
; Sequence 12, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent Ver. 2.0 - beta
; SEQ ID NO 12
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: (3)
; OTHER INFORMATION: Undetermined
US-08-913-430-12

Query Match 97.7%; Score 85; DB 13; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGXLQNSLLEEVWYAL 18
| | | | | | | | | | | | | | | | | |
Db 1 AGXLQNSLLEEVWYAL 18

RESULT 2
US-08-373-957-2
; Sequence 2, Application US/08373957
; GENERAL INFORMATION:
; APPLICANT: Wise, Kim S.

APPLICANT: McIntosh, Mark A.
TITLE OF INVENTION: DNA Sequences Coding for Mycoplasma
TITLE OF INVENTION: Hypopneumoniae Surface Antigens/
TITLE OF INVENTION: Corresponding Proteins and Use in Vaccines
TITLE OF INVENTION: and Diagnostic Procedures
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grace J. Fishel
STREET: 929 Fee Fee Road, Suite 100
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63043
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
COMPUTER: Hewlett-Packard Vectra 3.3
OPERATING SYSTEM: MS-DOS Version 5.1
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/373,957
FILING DATE: January 17, 1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fishel, Grace J.
REGISTRATION NUMBER: 25864
REFERENCE/DOCKET NUMBER: UVM 8141
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 878-0440
TELEFAX: (314) 275-7693
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 627 amino acid residues
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: Predicted amino acid sequence of complete
DESCRIPTION: 627 residues of the P65 lipoprotein, derived
DESCRIPTION: from the nucleic acid sequence
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: whole polypeptide
ORIGINAL SOURCE:
ORGANISM: Mycoplasma hypopneumoniae
STRAIN: J
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE: unicellular bacterium
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY: Genomic in Charon 4A, GEM12
CLONE: MhpJ25, MhpJ35, MhpJG35, p2J25,
CLONE: p2J25.1, p2J25.14, p2JG35.1, p2JG35.12,
CLONE: p2JG35.13, p2JG35.14
POSITION IN GENOME:
CHROMOSOME/SEGMENT: single chromosome
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: 627 amino acid sequence representing
NAME/KEY: complete sequence (including signal
NAME/KEY: sequence) of surface lipoprotein P65
LOCATION: entire derived coded sequence
IDENTIFICATION METHOD: clone identified by immunodetection of
IDENTIFICATION METHOD: protein product with antiserum specific for
IDENTIFICATION METHOD: P65; residue sequence deduced from nucleic

IDENTIFICATION METHOD: acid sequence
OTHER INFORMATION: immunogenic surface lipoprotein of no known
OTHER INFORMATION: function; C-terminus exposed on external
OTHER INFORMATION: surface of cell; N-terminal signal sequence
OTHER INFORMATION: (first 29 amino acid residues) cleaved during
OTHER INFORMATION: lipid modification process
PUBLICATION INFORMATION:
AUTHORS: Mary F. Kim, Maniieh B. Heidari, Susan J.
AUTHORS: Stull, Mark A. McIntosh, and Kim S. Wise
TITLE: Identification and Mapping of an
TITLE: Immunogenic Region of Mycoplasma
TITLE: hypopneumoniae p65 Surface Lipoprotein
TITLE: Expressed in Escherichia coli from a Cloned
TITLE: Genomic Fragment
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 8
PAGES: 2637-2643
DATE: August 1990
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: From 1 to 627
US-08-373-957-2
Query Match 73.6%; Score 64; DB 6; Length 627;
Best Local Similarity 83.3%; Pred. No. 0.0078;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AGXLOKNSLLEEVWYIAL 18
DB 28 AGCLQKNSLLSEVNYIAL 45
RESULT 3
US-09-150-857-6
Sequence 6, Application US/09150857
GENERAL INFORMATION:
APPLICANT: White, David
APPLICANT: Zhou, Jianghong
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: LEPTIN INDUCED GENES
FILE REFERENCE: 07334/093001
CURRENT APPLICATION NUMBER: US/09/150,857
CURRENT FILING DATE: 1998-09-10
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 400
TYPE: PRT
ORGANISM: Mus musculus
US-09-150-857-6
Query Match 49.4%; Score 43; DB 15; Length 400;
Best Local Similarity 61.5%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 4 LQKNSLLEEVWYIL 16
DB 365 LFKNSFLKEIFYL 377
RESULT 4
PCT-US99-16478-11
Sequence 11, Application PC/TUS9916478
GENERAL INFORMATION:
APPLICANT: E. I. du Pont de Nemours and Company
TITLE OF INVENTION: Plant Retinoblastoma-Related Cell Cycle Polypeptides
FILE REFERENCE: BB-1195
CURRENT APPLICATION NUMBER: PCT/US99/16478
CURRENT FILING DATE: 1999-07-21

EARLIER APPLICATION NUMBER: 60/093.767
EARLIER FILING DATE: July 22, 1998
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Microsoft Office 97
SEQ ID NO 11
LENGTH: 961
TYPE: PRT
ORGANISM: Nicotiana tabacum
PCT-US99-16478-11

Query Match 47.1%; Score 41; DB 1; Length 961;
Best Local Similarity 46.2%; Pred. No. 1.5e+02;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 GXLQNSLLEEVW 14
||:|:| ||:|:|
DB 471 GSLQSTSLMDNIW 483

RESULT 5
US-09-390-081-2
Sequence 2, Application US/09390081
GENERAL INFORMATION:
APPLICANT: David Malcolm Duckworth
APPLICANT: Joanne Rachel Evans
TITLE OF INVENTION: Novel Compounds
FILE REFERENCE: GP-30177
CURRENT APPLICATION NUMBER: US/09/390,081
CURRENT FILING DATE: 1999-09-03
EARLIER APPLICATION NUMBER: UK 9819405.3
EARLIER FILING DATE: 1998-09-04
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 642
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-390-081-2

Query Match 47.1%; Score 41; DB 17; Length 642;
Best Local Similarity 52.6%; Pred. No. 91;
Matches 10; Conservative 3; Mismatches 2; Indels 4; Gaps 1;

QY 4 LOKNSLLEE-----VWYLAL 18
||:|:|:| ||:|:|:|
DB 223 LORSSGMNETGVIVWYLAL 241

RESULT 6
US-09-390-081-4
Sequence 4, Application US/09390081
GENERAL INFORMATION:
APPLICANT: David Malcolm Duckworth
APPLICANT: Joanne Rachel Evans
TITLE OF INVENTION: Novel Compounds
FILE REFERENCE: GP-30177
CURRENT APPLICATION NUMBER: US/09/390,081
CURRENT FILING DATE: 1999-09-03
EARLIER APPLICATION NUMBER: UK 9819405.3
EARLIER FILING DATE: 1998-09-04
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 4
LENGTH: 642
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-390-081-4

Query Match 47.1%; Score 41; DB 17; Length 642;
Best Local Similarity 52.6%; Pred. No. 91;

Matches 10; Conservative 3; Mismatches 2; Indels 4; Gaps 1;
QY 4 LOKNSLLEE-----VWYLAL 18
||:|:|:| ||:|:|:|
DB 223 LORSSGMNETGVIVWYLAL 241

RESULT 7
US-08-827-356-5047
Sequence 5047, Application US/08827356
GENERAL INFORMATION:
APPLICANT: George H. Shimer, Jr.
APPLICANT: George H. Miller
APPLICANT: Roberta S. Hare
APPLICANT: Karen J. Shaw
TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS RELATED
TITLE OF INVENTION: COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 5574
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering-Plough Corporation
STREET: 2000 Galloping Hill Road
CITY: Kenilworth
STATE: New Jersey
COUNTRY: USA
ZIP: 07033-0530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/827,356
FILING DATE: 01-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,477
FILING DATE: 01-APR-1996
APPLICATION NUMBER: 60/016,743
FILING DATE: 02-MAY-1996
APPLICATION NUMBER: 60/020,016
FILING DATE: 14-JUN-1996
INFORMATION FOR SEQ ID NO: 5047:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...140
US-08-827-356-5047

Query Match 46.0%; Score 40; DB 12; Length 140;
Best Local Similarity 25.0%; Pred. No. 22;
Matches 4; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGXLQNSLLEEVWYL 16
||:|:|:| ||:|:|:|
DB 77 SGKHHNAIETIWFV 92

RESULT 8
US-09-082-969-2
Sequence 2, Application US/09082969
GENERAL INFORMATION:
APPLICANT: POPOFF, MICHEL Y.
APPLICANT: LE GUERN FELLOUS, MURIEL
TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE DETECTION OF
TITLE OF INVENTION: SALMONELLA
NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,969
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/586,272
; FILING DATE: 16-JAN-1996
; APPLICATION NUMBER: FR 9500410
; FILING DATE: 16-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-107-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-082-969-2

Query Match 46.0%; Score 40; DB 14; Length 553;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 LQKNSLLEEV 14
: |||:||||
DB 55 VSKNTLLDOV 65

RESULT 9
US-08-633-414-2
; Sequence 2, Application US/08633414
; GENERAL INFORMATION:
; APPLICANT: Sims, John E.
; APPLICANT: Virca, G. Duke
; APPLICANT: Bird, Timothy A.
; APPLICANT: Anderson, Dirk M.
; TITLE OF INVENTION: IL-1/TNF-(-ACTIVATED KINASE (ITAK),
; TITLE OF INVENTION: AND METHODS OF MAKING AND USING THE SAME
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,414
; FILING DATE: 10-JUN-1996

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 480052.418
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 979 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-414-2

Query Match 45.4%; Score 39.5; DB 10; Length 979;
Best Local Similarity 60.0%; Pred. No. 2.8e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

QY 5 QKNSLLEE---VWYL 16
: ||: ||| ||||
DB 143 QKDKLFEEMVWYL 157

RESULT 10
US-08-870-529-2
; Sequence 2, Application US/08870529
; GENERAL INFORMATION:
; APPLICANT: Sims, John E.
; APPLICANT: Virca, G. Duke
; APPLICANT: Bird, Timothy A.
; APPLICANT: Anderson, Dirk M.
; TITLE OF INVENTION: IL-1/TNF-(-ACTIVATED KINASE (ITAK),
; TITLE OF INVENTION: AND METHODS OF MAKING AND USING THE SAME
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,529
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 480052.418
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 979 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-870-529-2

Query Match 45.4%; Score 39.5; DB 12; Length 979;
Best Local Similarity 60.0%; Pred. No. 2.8e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

```

QY 5 QKNSLLEE---VWYL 16
   ||: ||| ||||
Db 143 QKDKLFEENVVWYL 157

RESULT 11
US-09-134-000-4234
; Sequence 4234, Application US/09134000A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS
; FILE REFERENCE: GTC-005
; CURRENT APPLICATION NUMBER: US/09/134,000A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 6810
; SEQ ID NO 4234
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000-4234

Query Match 44.8%; Score 39; DB 15; Length 483;
Best Local Similarity 46.7%; Pred. No. 1.5e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 LKNSLLEEVWYL 18
   ||: ||| ||||
Db 88 LDKRNLGSQFWYLAI 102

RESULT 12
US-60-122-952-5
; Sequence 5, Application US/60122952
; GENERAL INFORMATION:
; APPLICANT: Breinig, Sabine
; TITLE OF INVENTION: Phenol Induced Proteins of Thauera aromatica
; FILE REFERENCE: BC-1006-P1
; CURRENT APPLICATION NUMBER: US/60/122,952
; CURRENT FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Microsoft Word Version 97 SR-2
; SEQ ID NO 5
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Thauera aromatica
US-60-122-952-5

Query Match 44.8%; Score 39; DB 19; Length 612;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 7 NSLLEEVWYL 18
   ::: || ||||
Db 61 DTIWEAWYL 72

RESULT 13
US-60-138-676-545
; Sequence 545, Application US/60138676
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CLO00027
; CURRENT APPLICATION NUMBER: US/60/138,676
; CURRENT FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 609
; SOFTWARE: FastSeq for Windows Version 3.0

```

```

; SEQ ID NO 545
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Drosophila
US-60-138-676-545

Query Match 43.7%; Score 38; DB 19; Length 147;
Best Local Similarity 63.6%; Pred. No. 53;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LKNSLLEEVW 14
   ||: ||| ||||
Db 80 LEKLELLEELW 90

RESULT 14
US-60-140-803-613
; Sequence 613, Application US/60140803
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CLO00042
; CURRENT APPLICATION NUMBER: US/60/140,803
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 613
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Drosophila
US-60-140-803-613

Query Match 43.7%; Score 38; DB 19; Length 147;
Best Local Similarity 63.6%; Pred. No. 53;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LKNSLLEEVW 14
   ||: ||| ||||
Db 80 LEKLELLEELW 90

RESULT 15
US-60-143-993-1840
; Sequence 1840, Application US/60143993
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CLO00053
; CURRENT APPLICATION NUMBER: US/60/143,993
; CURRENT FILING DATE: 1999-07-15
; NUMBER OF SEQ ID NOS: 1906
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1840
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Drosophila
US-60-143-993-1840

Query Match 43.7%; Score 38; DB 19; Length 290;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LKNSLLEEVW 14
   ||: ||| ||||
Db 223 LEKLELLEELW 233

```

Search completed: November 13, 1999, 05:08:40
Job time: 10866 sec

A;Accession: S75860

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-179 <KAN>
A:Cross-references: EMBL:D90913; GB:AB001339; NID:g1653348; PID:d1019052; PID:g1653405
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Start codon: GTG

Query Match 47.1%; Score 41; DB 2; Length 179;
Best Local Similarity 61.5%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 4 LQNSLLEEVWYL 16
I I I I I I I I
DB 50 LASNGLLEAQWYL 62
RESULT 3
C70104
DNA topoisomerase IV (parC) homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
C:Accession: C70104
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: C70104
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-626 <KLE>
A:Cross-references: GB:AE001117; GB:AE000783; NID:g2687907; PID:g2687910; TIGR:BB0035
A:Experimental source: strain B31

Query Match 46.0%; Score 40; DB 2; Length 626;
Best Local Similarity 46.7%; Pred. No. 37;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
QY 4 LQNSLLEEVWYAL 18
I I I I I I I I
DB 329 LQSKILEKIFYKT 343
RESULT 4
S70817
invasion genes transcription activator hila - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Feb-1998
C:Accession: S70817; S70816
R:Bağaj, V.; Hwang, C.; Lee, C.A.
submitted to the EMBL Data Library, April 1995
A:Description: hila is a novel ompR/toxR family member that activates the expression of
A:Reference number: S70817
A:Accession: S70817
A:Molecule type: DNA
A:Residues: 1-553 <BAJ>
A:Cross-references: EMBL:U25352; NID:g1050873; PID:g806901
A:Experimental source: strain SL1344
A:Note: it is uncertain whether Met-1 or Met-23 is the initiator
R:Bağaj, V.; Hwang, C.; Lee, C.A.
Mol Microbiol 18, 715-727, 1995
A:Title: hila is a novel ompR/toxR family member that activates the expression of Salmon
A:Reference number: S70816; MUID:96414472
A:Accession: S70816
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 23-27,29-110 <BA2>
A:Cross-references: EMBL:U25352

A:Experimental source: strain SL1344
C:Genetics:
A:Gene: hila
C:Keywords: DNA binding; transcription regulation

Query Match 46.0%; Score 40; DB 2; Length 553;
Best Local Similarity 54.5%; Pred. No. 32;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 LQNSLLEEVW 14
I I I I I I I I
DB 55 VSKNTLLDQW 65

RESULT 5
S74742
exopolysaccharide export protein - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein sil10923
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
C:Accession: S74742
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S74742
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-756 <KAN>
A:Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017626; PID:g165196
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: epsB

Query Match 46.0%; Score 40; DB 2; Length 756;
Best Local Similarity 72.7%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 LQNSLLEEVW 14
I I I I I I I I
DB 112 LQSNILLEPVW 122

RESULT 6
S64386
pre-mRNA splicing protein PRP31 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein G4641; protein YGR091W
C:Species: Saccharomyces cerevisiae
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 06-Feb-1998
C:Accession: S64386; S64396; S59674; S65569
R:Wedler, H.; Scharfe, M.; Wedler, E.; Wambutt, R.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64356
A:Accession: S64386
A:Molecule type: DNA
A:Residues: 1-494 <WED>
A:Cross-references: EMBL:Z72876; NID:g1323134; PID:e243475; PID:g1323135; MIPS:YGR091
A:Experimental source: strain S288C
R:Hernandez, K.; Weber, N.; Wipfli, P.; Schmidheini, T.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64392
A:Accession: S64396
A:Molecule type: DNA
A:Residues: 1-494 <HER>
A:Cross-references: EMBL:Z72876; NID:g1323134; PID:e243475; PID:g1323135; MIPS:YGR091
A:Experimental source: strain S288C
R:Weidenhammer, E.M.; Singh, M.; Woolford, J.L.
submitted to the EMBL Data Library, July 1995

A:Description: The PRP31 gene encodes a novel protein required for pre-mRNA splicing in
A:Reference number: S59674
A:Accession: S59674
A:Molecule type: DNA
A:Residues: 1-360, 'TN', 363-364, 'R', 366, 'EL', 369, 'ENLGSTRKNSDYRMSDNKIGWNLA', 395-494 <WEI
A:Cross-references: EMBL:U31970; NID:g969100; PID:g969101
R:Weidenhammer, E.M.; Singh, M.; Ruiz-Noriega, M.; Woolford Jr., J.L.
Nucleic Acids Res. 24, 1164-1170, 1996
A:Title: The PRP31 gene encodes a novel protein required for pre-mRNA splicing in Saccha
A:Reference number: S65569
A:Accession: S65569
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-360, 'TN', 363-364, 'R', 366, 'EL', 369, 'ENLGSTRKNSDYRMSDNKIGWNLA', 395-494 <WEI
A:Cross-references: EMBL:U31970; NID:g969100; PID:g969101
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1995
C:Genetics:
A:Gene: SGD:PRP31
A:Cross-references: SGD:S0003323; MIPS:YGR091W
A:Map position: 7R
C:Function:
A:Description: required for pre-mRNA splicing

Query Match 46.0%; Score 40; DB 2; Length 494;
Best Local Similarity 53.8%; Pred. No. 28;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 LOKNSLLEEVWYL 16
| : |||| |
DB 198 LEANSILENLWKL 210

RESULT 7
Ysw1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YBR1125; Protein YBR148W
C:Species: Saccharomyces cerevisiae
A:Variety: Strain S288C
C:Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 06-Feb-1998
C:Accession: S46019; S48305
R:Entian, K.D.; Koetter, P.; Rose, M.; Becker, J.; Grey, M.; Li, Z.; Niegemann, E.; Sche
S.: Gruenbein, R.; Hedges, D.; Kiesel, P.; Korol, S.; Krems, B.; Proft, M.; Siegers, K.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S46013
A:Accession: S46019
A:Molecule type: DNA
A:Residues: 1-609 <ENT>
A:Cross-references: EMBL:U36017; NID:g536471; PID:g536472; MIPS:YBR148W
R:Schricker, R.; Ross-Macdonald, P.B.; Kandier, D.; Oberkofler, J.; Breitenbach, M.
submitted to the EMBL Data Library, August 1994
A:Description: Molecular characterization of the spore-specific gene ysw1 from Saccharom
A:Reference number: S48305
A:Accession: S48305
A:Molecule type: DNA
A:Residues: 1-57, 'NQ', 119-222, 'THR', 225-270, 'T', 272-364, 'M', 366-414, 'HE', 417-
A:Cross-references: EMBL:X80913; NID:g551671; PID:g551672
C:Genetics:
A:Gene: SGD:Ysw1
A:Cross-references: SGD:S0000352; MIPS:YBR148W
A:Map position: 2R

Query Match 45.4%; Score 39.5; DB 2; Length 609;
Best Local Similarity 60.0%; Pred. No. 43;
Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 2 GXLOKNSLLEEVWYL 16
| ||||| | : |
DB 28 GILQKNSLRLN-WFL 41

RESULT 8

A28171
phenylethanolamine N-methyltransferase (EC 2.1.1.28) - human
C:Species: Homo sapiens (man)
C:Date: 20-Jun-1989 #sequence_revision 20-Jun-1989 #text_change 24-Sep-1998
C:Accession: A28171; S10894; A28210
R:Kaneda, N.; Ichinose, H.; Kobayashi, K.; Oka, K.; Kishi, F.; Nakazawa, A.; Kurosawa
J. Biol. Chem. 263, 7672-7677, 1988
A:Title: Molecular cloning of cDNA and chromosomal assignment of the gene for human p
A:Reference number: A28171; MUID:88227966
A:Accession: A28171
A:Molecule type: mRNA
A:Residues: 1-282 <KAN>
A:Cross-references: GB:J03727; NID:g190141; PID:g190142
R:Sasaoka, T.; Kaneda, N.; Kurosawa, Y.; Fujita, K.; Nagatsu, T.
Neurochem. Int. 15, 555-565, 1989
A:Title: Structure of human phenylethanolamine N-methyltransferase gene: existence of
A:Reference number: S10894
A:Accession: S10894
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-282 <SAS>
A:Cross-references: EMBL:X52730; NID:g35560; PID:g296668
R:Baetge, E.E.; Behringer, R.R.; Messing, A.; Brinster, R.L.; Palmiter, R.D.
Proc. Natl. Acad. Sci. U.S.A. 85, 3648-3652, 1988
A:Title: Transgenic mice express the human phenylethanolamine N-methyltransferase gen
A:Reference number: A28210; MUID:88217959
A:Accession: A28210
A:Molecule type: mRNA
A:Residues: 1-168, 'AO', 171-282 <BAE>
A:Cross-references: GB:J03280; NID:g190143; PID:g387032
C:Genetics:
A:Gene: GDB:PNMT; PENT
A:Cross-references: GDB:I20271; OMIM:171190
A:Map position: 17pter-17qter
A:Introns: 68/1; 137/2
C:Superfamily: phenylethanolamine N-methyltransferase
C:Keywords: methyltransferase

Query Match 44.8%; Score 39; DB 2; Length 284;
Best Local Similarity 56.2%; Pred. No. 22;
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 GXLOKNSLLEEVWYL 17
| : |||||
DB 209 GHLLIGALEESWYLA 224

RESULT 9
A24313
phenylethanolamine N-methyltransferase (EC 2.1.1.28) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 29-Jan-1999
C:Accession: A24313
R:Baetge, E.E.; Suh, Y.H.; Joh, T.H.
Proc. Natl. Acad. Sci. U.S.A. 83, 5454-5458, 1986
A:Title: Complete nucleotide and deduced amino acid sequence of bovine phenylethanol
A:Reference number: A24313; MUID:86287277
A:Accession: A24313
A:Molecule type: mRNA
A:Residues: 1-284 <BAE>
A:Cross-references: GB:M14318; NID:g163553; PID:g163554
C:Superfamily: phenylethanolamine N-methyltransferase
C:Keywords: methyltransferase

Query Match 44.8%; Score 39; DB 2; Length 284;
Best Local Similarity 56.2%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 GXLOKNSLLEEVWYL 17
| : |||||
DB 209 GHLLIGALEESWYLA 224

A:Cross-references: EMBL:Z49437; NID:g1008347; PID:g1008348; MIPS:YJL162c

C:Genetics:

A:Map position: 10L

C:Superfamily: dnaJ amino-terminal homology

F:13-77/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 44.8%; Score 39; DB 2; Length 482;

Best Local Similarity 42.9%; Pred. No. 41;

Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGXLQKNSLLEEVW 14

||: ||: ||: ||:

Db 190 AGKMDTNSQFQIHW 203

RESULT 13

D71141

hypothetical protein PH0343 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998

C:Accession: D71141

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yanamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137

A:Accession: D71141

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-170 <KAW>

A:Cross-references: GB:AP000002; NID:g3236129; PID:g1030360; PID:g3256734

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH0343

Query Match 44.8%; Score 39; DB 2; Length 170;

Best Local Similarity 46.2%; Pred. No. 13;

Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 LQKNSLLEEVWYL 16

::: ||| ||:

Db 1 MKMNQLLEAFWI 13

RESULT 14

VCBEM7

major capsid protein - saimirine herpesvirus 1 (strain 11)

C:Species: saimirine herpesvirus 1

A:Note: host Saimiri sciureus (common squirrel monkey)

C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 05-Sep-1997

C:Accession: E36808

R:Albrecht, J.

submitted to the EMBL Data Library, January 1992

A:Description: Primary structure of the herpesvirus saimir genome.

A:Reference number: A36806

A:Accession: E36808

A:Molecule type: DNA

A:Residues: 1-1371 <ALB>

A:Cross-references: GB:X64346; NID:g60320; PID:g60346

R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.;

J. Virol. 66, 5047-5058, 1992

A:Title: Primary structure of the herpesvirus saimir genome.

A:Reference number: A37309; MUID:92333688

A:Contents: annotation; possible protein-coding frames

A:Note: neither amino acid nor nucleotide sequence is given

C:Genetics:

A:Gene: 25

C:Superfamily: varicella-zoster virus major capsid protein

C:Keywords: capsid protein

RESULT 10

I45962

phenylethanolamine N-methyltransferase - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Feb-1997

C:Accession: I45962

R:Batter, D.K.; D'Mello, S.R.; Turzai, L.M.; Hughes, H.B.

J. Neurosci. Res. 19, 367-376, 1988

A:Title: The complete nucleotide sequence and structure of the gene encoding bovine phen

A:Reference number: I45962; MUID:88245208

A:Accession: I45962

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-283 <BAT>

A:Cross-references: GB:M36706; NID:g163555; PID:g163556

C:Genetics:

A:Introns: 68/1; 137/2

C:Superfamily: phenylethanolamine N-methyltransferase

Query Match 44.8%; Score 39; DB 2; Length 283;

Best Local Similarity 56.2%; Pred. No. 23;

Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 GXLQKNSLLEEVWYLA 17

||: ||| ||| |||

Db 209 GHLLIGALEESWYLA 224

RESULT 11

A41481

virulence-associated protein virA - Salmonella typhimurium plasmid pYA426

C:Species: Salmonella typhimurium

C>Date: 10-Apr-1992 #sequence_revision 10-Apr-1992 #text_change 31-Oct-1997

C:Accession: A41481

R:Gulig, P.A.; Chiodo, V.A.

Infect. Immun. 58, 2651-2658, 1990

A:Title: Genetic and DNA sequence analysis of the Salmonella typhimurium virulence plasm

A:Reference number: A41481; MUID:90316693

A:Accession: A41481

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-241 <GUL>

A:Cross-references: GB:M64295; GB:M33927; NID:g154440; PID:g154442

C:Genetics:

A:Genome: plasmid

C:Superfamily: virulence-associated protein mkfa

Query Match 44.8%; Score 39; DB 2; Length 241;

Best Local Similarity 50.0%; Pred. No. 19;

Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 GXLQKNSLLEEVWY 15

||: ||| ||| |||

Db 223 GEMQRALEEPFY 236

RESULT 12

S56945

protein YJL162c - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein J0549

C:Species: Saccharomyces cerevisiae

C>Date: 05-May-1995 #sequence_revision 08-Sep-1995 #text_change 05-Dec-1997

C:Accession: S56945

R:Obermaier, B.; Piravandi, E.; Rinke, M.; Dondéy, H.

submitted to the Protein Sequence Database, September 1995

A:Reference number: S56937

A:Molecule type: DNA

A:Residues: 1-482 <OBE>

Search completed: November 13, 1999, 12:08:07
Job time: 2079 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:33:42 ; Search time 51.07 Seconds
(without alignments)
9.963 Million cell updates/sec

Title: US-08-913-430-12

Perfect score: 87

Sequence: 1 AGXLQKNSLLEEVWYLAL 18

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	52.9	836	1 DPOL_HPBDU	P03162 duck hepati
2	40	46.0	446	1 FAT_MOUSE	P70375 mus musculu
3	40	46.0	553	1 IAGA_SALTI	P43016 salmonella
4	40	46.0	553	1 IAGA_SALTY	P43015 salmonella
5	40	46.0	626	1 PARC_BORBU	O51066 borrelia bu
6	40	46.0	494	1 PR31_YEAST	P43704 saccharomyc
7	39.5	45.4	609	1 YSW1_YEAST	P38280 saccharomyc
8	39	44.8	283	1 PNMT_BOVIN	P10938 bos taurus
9	39	44.8	282	1 PNMT_HUMAN	P11086 homo sapien
10	39	44.8	295	1 PNMT_MOUSE	P40935 mus musculu
11	39	44.8	482	1 YJ02_YEAST	P45997 saccharomyc
12	38.5	44.3	1371	1 VCAP_HSVSA	Q00999 herpesvirus
13	38	43.7	285	1 PNMT_RAT	P10937 rattus norv
14	38	43.7	449	1 UF01_MANES	Q40284 manihot esc
15	38	43.7	394	1 UF06_MANES	Q40288 manihot esc
16	38	43.7	258	1 Y875_SYNY3	P73555 synchocyst
17	38	43.7	255	1 YC27_CYACA	P28257 cyanidium c
18	37.5	43.1	213	1 NCCN_ALCXX	Q44587 alcaligenes
19	37	42.5	3712	1 ACVS_CEPAC	P25464 cephalospor
20	37	42.5	747	1 AT12_HSVB	P28937 equine herp
21	37	42.5	1938	1 MYSD_CAEEL	P02567 caenorhabdi
22	37	42.5	497	1 SC14_YARLI	P45816 varrovia li
23	37	42.5	235	1 YDQ_Q_HAEIN	O57020 haemophilus
24	37	42.5	1888	1 YD72_SCHPO	O14207 schizosacch
25	37	42.5	530	1 YHG1_CLOPA	Q46189 clostridium
26	36.5	42.0	573	1 YKX3_CAEEL	P34280 caenorhabdi
27	36	41.4	745	1 ALD_HUMAN	P33897 homo sapien
28	36	41.4	736	1 ALD_MOUSE	P48410 mus musculu
29	36	41.4	1215	1 ATC6_YEAST	P39986 saccharomyc
30	36	41.4	416	1 CITG_HAEIN	P44458 haemophilus
31	36	41.4	349	1 END3_YEAST	P39013 saccharomyc
32	36	41.4	4451	1 GRSB_BACBR	P14688 bacillus br
33	36	41.4	82	1 IMPC_SALTY	P18643 salmonella
34	36	41.4	1127	1 MEM1_HUMAN	P55160 homo sapien
35	36	41.4	315	1 PSTC_HAEIN	P43191 haemophilus
36	36	41.4	527	1 RBF1_CANAL	Q00312 candida alb
37	36	41.4	183	1 RBS4_MESCR	Q08184 mesembryant
38	36	41.4	182	1 RBS5_MESCR	Q08185 mesembryant
39	36	41.4	186	1 RBS6_MESCR	Q08186 mesembryant
40	36	41.4	360	1 RECF_ACTPL	P24718 actinobacil
41	36	41.4	359	1 RECF_HAEIN	P43767 haemophilus
42	36	41.4	384	1 RECF_MYCSM	P50916 mycobacteri
43	36	41.4	479	1 SAHH_PLAF7	P50250 plasmodium

ALIGNMENTS

```
RESULT 1
DPOL_HPBDU STANDARD; PRT; 836 AA.
AC P03162:
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DNA POLYMERASE (EC 2.7.7.7).
GN P.
OS DUCK HEPATITIS B VIRUS (DHBV).
OC VIRUSES; RETROID VIRUSES; HEPADNAVIRIDAE; AVIHEPADNAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84138772.
RA MANDART E., KAY A., GALIBERT F.;
RT "Nucleotide sequence of a cloned duck hepatitis B virus genome:
RT comparison with woodchuck and human hepatitis B virus sequences.";
RL J. VIROL. 49:782-792(1984).
RN [2]
RP SEQUENCE OF 794-836 FROM N.A.
RX MEDLINE: 84216498.
RA MOLNAR-KIMBER K.L., SUMMERS J.W., MASON W.S.;
RT "Mapping of the cohesive overlap of duck hepatitis B virus DNA and of
RT the site of initiation of reverse transcription.";
RL J. VIROL. 51:181-191(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K01834; G325433; -
DR PIR; A00710; JDVLD.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF00242; DNA_pol_viral_N; 1.
DR PFAM; PF00336; DNA_pol_viral_C; 1.
DR PFAM; PF00336; DNA_pol_viral_C; 1.
KW TRANSFERASE; DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.
SQ SEQUENCE 836 AA; 95275 MW; B2784CCE CRC32;
-----
Query Match 52.9%; Score 46; DB 1; Length 836;
Best Local Similarity 61.1%; Pred. No. 3.1;
Matches 11; Conservative 1; Mismatches 4; Indels 2; Gaps 1:
QY 1 AGXLQKNSLLEEVWYL 16
Db 409 SGASDKNSPLEENWYL 426
-----
RESULT 2
FAT_MOUSE STANDARD; PRT; 446 AA.
ID FA7_MOUSE
AC P70375:
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE COAGULATION FACTOR VII PRECURSOR (EC 3.4.21.21).
GN F7 OR CF7.
OS MUS MUSCULUS (MOUSE).
OC EURARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
```

RX MEDLINE; 97127167.
RA IDUSOGIE E., ROSEN E.D., CARMELIET P., COLLEN D., CASTELLINO F.J.;
RT "Nucleotide structure and characterization of the murine blood
RT coagulation factor VII gene";
RL THROMB. HAEMOST. 76:957-964(1996).
CC -1- FUNCTION: CIRCULATES IN THE BLOOD IN A ZYMOGEN FORM. FACTOR VII IS
CC CONVERTED TO FACTOR VIIA BY FACTOR XA, FACTOR XIIIA, FACTOR IXA, OR
CC THROMBIN BY MINOR PROTEOLYSIS. IN THE PRESENCE OF TISSUE FACTOR
CC AND CALCIUM IONS, FACTOR VIIA THEN CONVERTS FACTOR X TO FACTOR XA
CC BY LIMITED PROTEOLYSIS. FACTOR VIIA WILL ALSO CONVERT FACTOR IX TO
CC FACTOR IXA IN THE PRESENCE OF TISSUE FACTOR AND CALCIUM (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: HYDROLYSES ONE ARG-ILE BOND IN FACTOR X TO
CC FORM FACTOR XA.
CC -1- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND A HEAVY CHAIN LINKED
CC BY A DISULFIDE BOND (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: PLASMA.
CC -1- PTM: THE VITAMIN K-DEPENDENT, ENZYMOLOGIC CARBOXYLATION OF SOME
CC GLUTAMIC ACID RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND
CC CALCIUM (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U66079; G151322; -
DR MGD; MGI:103325; CF7.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PS01187; EGF_CA; 1.
DR PFAM; PF00008; EGF; 1.
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00594; gla; 1.
DR HSP; P00740; IIXA.
KW HYDROLASE; SERINE PROTEASE; BLOOD COAGULATION; ZYMOGEN; GLYCOPROTEIN;
KW LIVER; PLASMA; VITAMIN K; CALCIUM-BINDING; GAMMA-CARBOXYGLUTAMIC ACID;
KW EGF-LIKE DOMAIN; REPEAT; SIGNAL; HYDROXYLATION.
FT SIGNAL 1 ?
FT PROPEP 1 ?
FT CHAIN 42 193
FT CHAIN 194 446
FT DOMAIN 47 76
FT DOMAIN 87 123
FT DOMAIN 128 169
FT DOMAIN 194 446
FT SITE 193 194
FT ACT_SITE 234 234
FT ACT_SITE 283 283
FT ACT_SITE 385 385
FT BINDING 379 379
FT BINDING 58 63
FT DISULFID 91 102
FT DISULFID 96 111
FT DISULFID 113 122
FT DISULFID 132 143
FT DISULFID 139 153
FT DISULFID 155 168
FT DISULFID 176 200
FT DISULFID 200 205
FT DISULFID 219 235
FT DISULFID 351 370

FT DISULFID 381 409 BY SIMILARITY.
FT MOD_RES 47 47 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 48 48 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 55 55 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 57 57 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 60 60 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 61 61 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 66 66 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 67 67 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 70 70 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 76 76 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 104 104 HYDROXYLATION (BY SIMILARITY).
FT CARBOHYD 186 186 POTENTIAL.
FT CARBOHYD 244 244 POTENTIAL.
SQ SEQUENCE 446 AA; 50276 MW; 21169FBB CRC32;

Query Match 46.0%; Score 40; DB 1; Length 446;
Best Local Similarity 87.5%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 NSLLEEVW 14
DB 43 NSLLEELW 50
|||||:|

RESULT 3
ID IAGA_SALTI STANDARD; PRT; 553 AA.
AC P43016;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE INVASION PROTEIN IAGA.
GN IAGA.
OS SALMONELLA TYPHI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TY2;
RC MEDLINE; 95273801.
RA MIRAS I., HERMANT D., ARRICAU N., POPOFF M.Y.;
RT "Nucleotide sequence of iaga and lagB genes involved in invasion of
RT HeLa cells by Salmonella enterica subsp. enterica ser. Typhi";
RL RES. MICROBIOL. 146:17-20(1995).
CC -1- FUNCTION: ACTIVATES THE EXPRESSION OF INVASION GENES.
CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X80892; G791140; -
DR PFAM; PF00486; trans_reg_C; 1.
KW SENSORY TRANSDUCTION; TRANSCRIPTION REGULATION; ACTIVATOR;
KW DNA-BINDING; VIRULENCE; PHOSPHORYLATION.
FT MOD_RES 62 62 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 553 AA; 63026 MW; BB62B37C CRC32;

Query Match 46.0%; Score 40; DB 1; Length 553;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 LKNSLLEEVW 14
DB 55 VSKNTLLDQW 65
|||||:|

```

RESULT 4
IAGA_SALTY
ID IAGA_SALTY STANDARD; PRT; 553 AA.
AC F43015;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE INVASION PROTEIN IAGA (PROTEIN HILA).
GN IAGA OR HILA.
OS SALMONELLA TYPHIMURIUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-S11344;
RX MEDLINE; 96414472.
RA BAJAJ V., HWANG C., LEE C.A.;
RT "hila" is a novel ompR/toxR family member that activates the
RT expression of Salmonella typhimurium invasion genes.";
RL MOL. MICROBIOL. 18:715-727(1995).
CC -!- FUNCTION: ACTIVATES THE EXPRESSION OF INVASION GENES.
CC -!- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U25352; G806901;
DR STYGENE; SG10537; IAGA.
DR PFAM; PF00486; trans_reg_C; 1.
DR HSSP; P03025; 10DD.
KW SENSORY TRANSDUCTION; TRANSCRIPTION REGULATION; ACTIVATOR;
KW DNA-BINDING; VIRULENCE; PHOSPHORYLATION.
FT MOD_RES 62 62
FT PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 553 AA; 63027 MW; FD7F9D4B CRC32;

Query Match 46.0%; Score 40; DB 1; Length 553;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 LQKNSLLEEVW 14
Db 55 VSKNTLLDQVW 65

RESULT 5
PARC_BORBU
ID PARC_BORBU STANDARD; PRT; 626 AA.
AC O51066;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TOPOISOMERASE IV SUBUNIT A (EC 5.99.1.1).
GN PARC OR BB0035.
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-ATCC 35210 / B31;
RX MEDLINE; 98065943.
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,

```

```

RA UTTERBACK T., WATTHEY L., McDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi.";
RL NATURE 390:580-586(1997).
CC -!- FUNCTION: TOPOISOMERASE IV IS ESSENTIAL FOR CHROMOSOME
CC SEGREGATION. IT HAS RELAXATION OF SUPERCOILED DNA ACTIVITY.
CC PERFORMS THE DECATENATION EVENTS REQUIRED DURING THE REPLICATION
CC OF A CIRCULAR DNA MOLECULE (BY SIMILARITY).
CC -!- SUBUNIT: COMPOSED OF TWO SUBUNITS: PARC AND PARE.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (BY SIMILARITY).
CC -!- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001117; G2687910;
DR TIGR; BB0035;
KW TOPOISOMERASE; ISOMERASE; DNA-BINDING.
FT ACT_SITE 105 105
FT DNA_CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 626 AA; 72041 MW; 295E2871 CRC32;

Query Match 46.0%; Score 40; DB 1; Length 626;
Best Local Similarity 46.7%; Pred. No. 24;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 LQKNSLLEEVWYIAL 18
Db 329 LQKSKILEKIFYKTL 343

RESULT 6
PR31_YEAST
ID PR31_YEAST STANDARD; PRT; 494 AA.
AC P49704;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PRE-MRNA SPLICING FACTOR PRP31.
GN PRP31 OR YGR091W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 96184869.
RA WEIDENHAMMER E.M., SINGH M., RUIZ-NORIEGA M., WOOLFORD J.L. JR.;
RT "The PRP31 gene encodes a novel protein required for pre-mRNA
RT splicing in Saccharomyces cerevisiae.";
RL NUCLEIC ACIDS RES. 24:1164-1170(1996).
RN [2]
RN SEQUENCE FROM N.A.
RA WEDLER H., SCHARFE M., WEDLER E., WAMBUTT R.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- FUNCTION: REQUIRED FOR PRE-MRNA SPLICING.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31970; G969101;

```

DR EMBL; Z72876; E243475; -
 DR SGD; L0002947; PRP31.
 KW MRNA PROCESSING; MRNA SPLICING; NUCLEAR PROTEIN.
 FT CONFLICT 361 394 TPKRKLGENLSTKNSYRMSDNKYKIGNLA ->
 FT DQPKKRRAGKRRKRIKFKRLSHVRQLQNRMEFG (IN
 FT REF. 2).
 SQ SEQUENCE 494 AA; 55967 MW; 6A60B5F5 CRC32;

Query Match 46.08; Score 40; DB 1; Length 494;
 Best Local Similarity 53.88; Pred. No. 19;
 Matches 7; Conservative 3; Mismatches 0; Gaps 0;

QY 4 LOKNSLLEEVWYL 16
 I: ||||| : |
 DB 198 LEANSILENLWL 210

RESULT 7
 YSW1 YEAST
 ID YSW1 YEAST STANDARD; PRT; 609 AA.
 AC P38280;
 DT 01-OCT-1994 (REL. 30, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE SPORE-SPECIFIC PROTEIN YSW1.
 GN YSW1 OR YBR148W OR YBR1125.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMICETES; SACCHAROMYCETALES;
 CC SACCHAROMYCETACEAE; SACCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SK1;
 RA SCHRICKER R., ROSS-MACDONALD P.B., KANDLER D., OBERKOFER J.,
 RA BREITENBACH M.;
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA ENTIAN K.-D., KOETTER P., ROSE M., BECKER J., GREY M., LI Z.,
 RA NIEGEMANN E., SCHENK-GROENINGER R., SERVOS J., WEHNER E.,
 RA WOLTER R., BRENDLE M., BAUER J., BRAUN H., DERN K., DUBSTERHUS S.,
 RA GRUENBEIN R., HEDGES D., KIESAU P., KOROL S., KREMS B., PROFT M.,
 RA SIEGERS K., BAUR A., BOLES E., MIOSGA T.,
 RA SCHAAFF-GERSTENSCHLAGER I., ZIMMERMANN F.K.;
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z36017; G536472; -
 DR EMBL; X80913; G551672; -
 DR PIR; S46019; S46019.
 DR SGD; L0002533; YSW1.
 FT CONFLICT 58 58 D -> N (IN REF. 1).
 FT CONFLICT 117 118 KE -> NQ (IN REF. 1).
 FT CONFLICT 223 224 TP -> IHR (IN REF. 1).
 FT CONFLICT 271 271 A -> T (IN REF. 1).
 FT CONFLICT 365 365 L -> M (IN REF. 1).
 FT CONFLICT 415 416 QQ -> HE (IN REF. 1).
 FT CONFLICT 460 460 D -> N (IN REF. 1).
 FT CONFLICT 481 482 LS -> SR (IN REF. 1).
 FT CONFLICT 488 488 HNYVND -> PMFMAMI (IN REF. 1).
 FT CONFLICT 495 609 MISSING (IN REF. 1).
 SQ SEQUENCE 609 AA; 70174 MW; 133092DF CRC32;

Query Match 45.48; Score 39.5; DB 1; Length 609;

Best Local Similarity 60.08; Pred. No. 29;
 Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;
 QY 2 GXLOKNSLLEEVWYL 16
 I: ||||| : |
 DB 28 GILQKNSTLRN-WFL 41

RESULT 8
 PNMT_BOVIN
 ID PNMT_BOVIN STANDARD; PRT; 283 AA.
 AC P10938;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE PHENYLETHANOLAMINE N-METHYLTRANSFERASE (EC 2.1.1.28) (PNMTASE)
 DE (NORADRENALINE N-METHYLTRANSFERASE).
 GN PNMT.

OS BOS TAURUS (BOVINE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 88245208.
 RA BATTER D.K., D'ELLO S.R., TURZAI L.M., HUGHES H.B. III, GIOIO A.E.,
 RA KAPLAN B.B.;
 RT "The complete nucleotide sequence and structure of the gene encoding
 RT bovine phenylethanolamine N-methyltransferase.";
 RL J. NEUROSCI. RES. 19:367-376(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 86287277.
 RA BAETGE E.E., SUH Y.H., JOH T.H.;
 RT "Complete nucleotide and deduced amino acid sequence of bovine
 RT phenylethanolamine N-methyltransferase: partial amino acid homology
 RT with rat tyrosine hydroxylase.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:5454-5458(1986).
 RN [3]
 RP PARTIAL SEQUENCE.
 RX MEDLINE; 88245209.
 RA WEISBERG E.P., BATTER D.K., BROWN W.E., KAPLAN B.B.;
 RT "Purification and partial amino acid sequence of bovine adrenal
 RT phenylethanolamine N-methyltransferase: a comparison of nucleic acid
 RT and protein sequence data.";
 RL J. NEUROSCI. RES. 19:377-382(1988).
 RN [4]
 RP PARTIAL SEQUENCE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M36706; G163556; -
 DR EMBL; M14318; G163554; -
 DR FIR; A24313; A24313.
 DR PROSITE; PS01100; NNMT_PNMT_TENT; 1.
 DR PFAM; PF01234; NNMT_PNMT_TENT; 1.

KW TRANSFERASE; METHYLTRANSFERASE; CATECHOLAMINE BIOSYNTHESIS.
FT CONFLICT 246 248 TAT -> SAM (IN REF. 2).
FT CONFLICT 255 RTMPFAHLQTDVDDKGIFFTR ->
FT CONFLICT 255 LHLHACPPSDRCRCOGLHLHL (IN REF. 2).
SQ SEQUENCE 283 AA; 30918 MW; 2C35554F CRC32;
Query Match 44.8%; Score 39; DB 1; Length 283;
Best Local Similarity 56.2%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 GXLOKNSLLEEVWYLA 17
DB 209 GHLLIGALEESWYLA 224
RESULT 9
PNMT_HUMAN
ID PNMT_HUMAN STANDARD; PRT; 282 AA.
AC P11086;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PHENYLETHANOLAMINE N-METHYLTRANSFERASE (EC 2.1.1.28) (PNMTASE)
DE (NORADRENALINE N-METHYLTRANSFERASE).
GN PNMT OR PENT.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88227966.
RA KANEDA N., ICHINOSE H., KOBAYASHI K., OKA K., KISHI F., NAKAZAWA A.,
RA KUROSAWA Y., FUJITA K., NAGATSU T.;
RT "Molecular cloning of cDNA and chromosomal assignment of the gene for
RT human phenylethanolamine N-methyltransferase, the enzyme for
RT epinephrine biosynthesis.";
RL J. BIOL. CHEM. 263:7672-7677(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA SASAKA T., KANEDA N., KUROSAWA Y., FUJITA K., NAGATSU T.;
RT "Structure of human phenylethanolamine N-methyltransferase gene:
RT existence of two types of mRNA with different transcription initiation
RT sites.";
RL NEUROCHEM. INT. 15:555-565(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88217959.
RA BAETGE E.E., BEHRINGER R.R., MESSING A., BRINSTER R.L., PALMITER R.D.;
RT "Transgenic mice express the human phenylethanolamine N-
RT methyltransferase gene in adrenal medulla and retina.";
RL PROC. NATL. ACAD. SCI. U.S.A. 85:3648-3652(1988).
CC -!- FUNCTION: CONVERTS NORADRENALINE TO ADRENALINE.
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + PHENYLETHANOLAMINE
CC = S-ADENOSYL-L-HOMOCYSTEINE + N-METHYLPHENYLETHANOLAMINE.
CC -!- PATHWAY: LAST STEP IN CATECHOLAMINE BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS EC 2.1.1.1,
CC EC 2.1.1.28 AND EC 2.1.1.96.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J03727; G190142; -
CC EMBL; X52730; G296668; -
CC EMBL; J03280; G387032; -
CC PIR; A28171; A28171.
CC PIR; S10894; S10894.

DR MIM; 171190; -
DR PROSITE; PS01100; NNMT_PNMT_TMT; 1.
DR PFAM; PF01234; NNMT_PNMT_TMT; 1.
KW TRANSFERASE; METHYLTRANSFERASE; CATECHOLAMINE BIOSYNTHESIS.
FT CONFLICT 169 170 SP -> AQ (IN REF. 3).
SQ SEQUENCE 282 AA; 30855 MW; 92649965 CRC32;
Query Match 44.8%; Score 39; DB 1; Length 282;
Best Local Similarity 56.2%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 GXLOKNSLLEEVWYLA 17
DB 209 GHLLIGALEESWYLA 224
RESULT 10
PNMT_MOUSE
ID PNMT_MOUSE STANDARD; PRT; 295 AA.
AC P40935;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PHENYLETHANOLAMINE N-METHYLTRANSFERASE (EC 2.1.1.28) (PNMTASE)
DE (NORADRENALINE N-METHYLTRANSFERASE).
GN PNMT.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA QUATRE C.J., HOYLE G.W., FROELICH G.J., FINDLEY S.D., BAETGE E.E.,
RA BEHRINGER R.R., PALMITER R.D.;
RL SUBMITTED (MAY-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2J;
RX MEDLINE; 92326574.
RA MORITA S., KOBAYASHI K., HIDAKA H., NAGATSU T.;
RT "Organization and complete nucleotide sequence of the gene encoding
RT mouse phenylethanolamine N-methyltransferase.";
RL BRAIN RES. MOL. BRAIN RES. 13:313-319(1992).
CC -!- FUNCTION: CONVERTS NORADRENALINE TO ADRENALINE.
CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + PHENYLETHANOLAMINE
CC = S-ADENOSYL-L-HOMOCYSTEINE + N-METHYLPHENYLETHANOLAMINE.
CC -!- PATHWAY: LAST STEP IN CATECHOLAMINE BIOSYNTHESIS.
CC -!- TISSUE SPECIFICITY: BRAIN (PONS AND MEDULLA OBLONGATA), ADRENAL
CC GLAND AND RETINA.
CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS EC 2.1.1.1,
CC EC 2.1.1.28 AND EC 2.1.1.96.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L12687; G293768; -
CC MGD; MGI:97724; PNMT.
DR PROSITE; PS01100; NNMT_PNMT_TMT; 1.
DR PFAM; PF01234; NNMT_PNMT_TMT; 1.
KW TRANSFERASE; METHYLTRANSFERASE; CATECHOLAMINE BIOSYNTHESIS; REPEAT.
FT DOMAIN 4 20 2 X 7 AA REPEATS OF G-S-D-[LP]-K-H-A.
FT REPEAT 4 10 1.
FT REPEAT 14 20 2.
SQ SEQUENCE 295 AA; 32544 MW; B3C0CEDC CRC32;
Query Match 44.8%; Score 39; DB 1; Length 295;
Best Local Similarity 56.2%; Pred. No. 16;

RT "Cloning of the rat adrenal medullary phenylethanolamine-N-methyltransferase.";
 RL NUCLEIC ACIDS RES. 17:2125-2125(1989).
 CC -!- FUNCTION: CONVERTS NORADRENALINE TO ADRENALINE.
 CC -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + PHENYLETHANOLAMINE
 CC = S-ADENOSYL-L-HOMOCYSTEINE + N-METHYLPHENYLETHANOLAMINE.
 CC -!- PATHWAY: LAST STEP IN CATECHOLAMINE BIOSYNTHESIS.
 CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS EC 2.1.1.1,
 CC EC 2.1.1.28 AND EC 2.1.1.96.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X75333; G414187;
 DR EMBL: U11694; G607861;
 DR EMBL: U11275; G607861; JOINED.
 DR EMBL: X14211; G56944;
 DR PIR: A60060; A60060.
 DR PIR: S03614; S03614.
 DR PROSITE: PS01100; NNMT_PNMT_TMT; 1.
 DR PFAM: PF01234; NNMT_PNMT_TMT; 1.
 KW TRANSFERASE; METHYLTRANSFERASE; CATECHOLAMINE BIOSYNTHESIS.
 FT CONFLICT 16 17 DS -> LA (IN REF. 3).
 FT CONFLICT 52 52 D -> H (IN REF. 4).
 FT CONFLICT 204 204 T -> A (IN REF. 4).
 FT CONFLICT 211 211 N -> H (IN REF. 2 AND 3).
 FT CONFLICT 214 214 F -> L (IN REF. 2 AND 3).
 FT CONFLICT 285 285 V -> A (IN REF. 2).
 SQ SEQUENCE 285 AA; 31670 MW; B8445DF6 CRC32;

Query Match 43.7%; Score 38; DB 1; Length 285;
 Best Local Similarity 87.5%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 LEEVWYLA 17
 ||| ||||
 DB 218 LEEWYLA 225

RESULT 14
 UF06_MANES STANDARD; PRT; 449 AA.
 AC Q40284;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 1 (EC 2.4.1.91) (UDP-GLUCOSE
 DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 1).
 GN GT1 OR UGT73AL.
 OS MANIHOT ESCULENTA (CASSAVA) (MANIOC).
 CC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 CC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 CC EUPHORBIALES; EUPHORBIACEAE; MANIHOT.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COTYLEDON;
 RX MEDLINE; 95201291.
 RA HUGHES J., HUGHES M.A.;
 RT "Multiple secondary plant product UDP-glucose glucosyltransferase
 RT genes expressed in cassava (Manihot esculenta crantz) cotyledons.";
 RL DNA SEQ. 5:41-49(1994).
 CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
 CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
 CC PIGMENTS (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL - UDP + FLAVONOL
 CC 3-O-D-GLUCOSIDE.
 CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE

CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN COTYLEDONS AND ROOTS, BUT NOT IN
 CC LEAVES.
 CC -!- DEVELOPMENTAL STAGE: MAXIMUM EXPRESSION IN COTYLEDONS THAT JUST
 CC EMERGED FROM THE SEED COAT. LOW LEVELS IN HYPOCOTYLS AND
 CC INCREASING LEVELS IN ROOTS THROUGHOUT THIS PERIOD OF DEVELOPMENT.
 CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X77459; G453246;
 DR PROSITE: PS00375; UDPGT; 1.
 DR PFAM: PF00201; UDPGT; 1.
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; MULTIGENE FAMILY.
 SQ SEQUENCE 449 AA; 50282 MW; ECA546CF CRC32;

 Query Match 43.7%; Score 38; DB 1; Length 449;
 Best Local Similarity 55.6%; Pred. No. 38;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 7 NSLLEEVWY 15
 ||:||||:
 DB 346 NSILESINF 354

 RESULT 15
 UF06_MANES STANDARD; PRT; 394 AA.
 ID Q40288;
 AC 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE FLAVONOL 3-O-GLUCOSYLTRANSFERASE 6 (EC 2.4.1.91) (UDP-GLUCOSE
 DE FLAVONOID 3-O-GLUCOSYLTRANSFERASE 6) (FRAGMENT).
 GN GT6 OR UGT73A6.
 OS MANIHOT ESCULENTA (CASSAVA) (MANIOC).
 CC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 CC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 CC EUPHORBIALES; EUPHORBIACEAE; MANIHOT.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COTYLEDON;
 RX MEDLINE; 95201291.
 RA HUGHES J., HUGHES M.A.;
 RT "Multiple secondary plant product UDP-glucose glucosyltransferase
 RT genes expressed in cassava (Manihot esculenta crantz) cotyledons.";
 RL DNA SEQ. 5:41-49(1994).
 CC -!- FUNCTION: IN THE PRESENCE OF OTHER NECESSARY COLOR FACTORS, THIS
 CC GLYCOSYLATION REACTION ALLOWS THE ACCUMULATION OF ANTHOCYANIN
 CC PIGMENTS (BY SIMILARITY). MAY BE INVOLVED IN GLYCOSYLATION OF
 CC INSTABLE CYANOHYDRINS TO PRODUCE STABLE CYANOGLUCOSIDES.
 CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + A FLAVONOL - UDP + FLAVONOL
 CC 3-O-D-GLUCOSIDE.
 CC -!- PATHWAY: ONE OF THE ESSENTIAL AND TERMINAL STEPS IN THE
 CC ANTHOCYANIN BIOSYNTHETIC PATHWAY.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN COTYLEDONS AND LEAVES.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY IN DEVELOPING COTYLEDONS
 CC (FROM EMERGING GREEN COTYLEDONS TO APPROXIMATELY 10 DAYS OLD);
 CC VERY LOW LEVELS IN HYPOCOTYLS AND NO MEASURABLE EXPRESSION IN
 CC ROOTS.
 CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X77463; G453251; -
 DR PROSITE; PS00375; UDPGT; 1.
 DR PFAM; PF00201; UDPGT; 1.
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; MULTIGENE FAMILY.
 FT NON_TER 1
 SQ SEQUENCE 394 AA; 44264 MW; 5657C860 CRC32;

Query Match 43.7%; Score 38; DB 1; Length 394;
 Best Local Similarity 55.6%; Pred. No. 32;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 7 NSLLEEVVY 15
 ||||:|:
 Db 288 NSILESIWF 296

Search completed: November 13, 1999, 10:33:44
 Job time: 5194 sec

GenCore version 4.5

Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:37 ; Search time 139.86 Seconds
(without alignments)
7.921 Million cell updates/sec

Title: US-08-913-430-12

Perfect score: 87
Sequence: 1 AGXLQKNSLLEEVWYAL 18

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPTREMBL_10.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64	73.6	627	2	O30643 mycoplasma
2	64	73.6	625	2	O30704 mycoplasma
3	46	52.9	836	12	O71306 duck hepati
4	46	52.9	786	12	O03765 duck hepati
5	46	52.9	836	12	O66400 duck hepati
6	45	51.7	558	5	O45151 caenorhabdi
7	43	49.4	413	11	O921M3 mus musculu
8	42	48.3	494	5	P91515 caenorhabdi
9	42	48.3	3085	5	O97283 plasmodium
10	42	48.3	1542	5	O97299 plasmodium
11	41	47.1	1779	2	P74225 synechocyst
12	41	47.1	3512	5	O62524 chironomus
13	40	46.0	756	2	P72877 synechocyst
14	40	46.0	193	3	O94520 schizosacch
15	40	46.0	252	5	O01437 caenorhabdi
16	40	46.0	291	10	O04267 arabidopsis
17	40	46.0	446	11	O61109 mus musculu
18	39.5	45.4	559	5	O94815 taenia sagi
19	39	44.8	170	1	O58081 pyrococcus
20	39	44.8	241	2	O04251 salmonella
21	39	44.8	1130	2	O46545 bacteroides
22	39	44.8	306	2	O51855 pseudomonas
23	39	44.8	1275	4	O15057 homo sapien
24	39	44.8	1455	4	O92497 homo sapien
25	39	44.8	1455	4	O15360 homo sapien
26	39	44.8	460	5	P91467 caenorhabdi
27	39	44.8	606	5	O44456 caenorhabdi
28	39	44.8	1236	5	O44875 caenorhabdi
29	38.5	44.3	821	5	O09396 caenorhabdi

30	38.5	44.3	105	12	Q69332 salmirine
31	38	43.7	306	2	P97091 pseudomonas
32	38	43.7	161	2	O67457 aquifex aeo
33	38	43.7	227	2	O46140 clostridium
34	38	43.7	293	2	O50727 borrelia bu
35	38	43.7	306	5	Q23084 caenorhabdi
36	38	43.7	1076	5	O17904 caenorhabdi
37	38	43.7	311	5	O60987 dictyosteli
38	38	43.7	759	12	Q67296 influenza a
39	38	43.7	759	12	Q67297 influenza a
40	37.5	43.1	189	2	O84213 chlamydia t
41	37	42.5	705	2	Q50468 mycobacteri
42	37	42.5	705	2	P96283 mycobacteri
43	37	42.5	248	3	O74926 schizosacch
44	37	42.5	459	5	P91466 caenorhabdi
45	37	42.5	5037	13	Q91313 rana catesb

ALIGNMENTS

RESULT 1					
O30643					
ID O30643	PRELIMINARY;	PRT;	627 AA.		
AC					
DT 01-JAN-1998 (TREMREL. 05, Created)					
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)					
DT 01-MAY-1999 (TREMREL. 10, Last annotation update)					
DE PROLIPOPROTEIN P65 PRECURSOR.					
OS Mycoplasma hyopneumoniae.					
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;					
OC Mycoplasmataceae; Mycoplasma.					
RN [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN-J, ATCC25334;					
RA HEIDARI M.B., KIM M.F., MCINTOSH M.A., WISE K.W.;					
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.					
DR EMBL; AF013714; AAB67173.1; -					
DR PFAM; PF00657; Lipase_GDSL; 1.					
KW Signal; Lipoprotein.					
FT SIGNAL 1 29	POTENTIAL.				
FT CHAIN 30 627	LIPOPROTEIN P65.				
SQ SEQUENCE 627 AA; 71016 MW; 10C5A2A5 CRC32;					
Query Match	73.6%;	Score 64;	DB 2;	Length 627;	
Best Local Similarity	83.3%;	Pred. No. 0.005;	Mismatches 0;	Indels 0;	Gaps 0;
Matches 15;	Conservative 0;				
QY 1 AGXLQKNSLLEEVWYAL 18					
Db 28 AGCLQKNSLSEVNYAL 45					
II IIIIIII IIIII					
RESULT 2					
O30704					
ID O30704	PRELIMINARY;	PRT;	625 AA.		
AC					
DT 01-JAN-1998 (TREMREL. 05, Created)					
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)					
DT 01-MAY-1999 (TREMREL. 10, Last annotation update)					
DE SURFACE LIPOPROTEIN PRECURSOR.					
OS Mycoplasma hyopneumoniae.					
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;					
OC Mycoplasmataceae; Mycoplasma.					
RN [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN-232A;					
RA HSU T., MINION F.C.;					
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.					
DR EMBL; AF015665; AAB70214.1; -					
DR PFAM; PF00657; Lipase_GDSL; 1.					
KW Signal; Lipoprotein.					

FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 625 MATURE SURFACE LIPOPROTEIN.
SQ SEQUENCE 625 AA; 70914 MW; C109763D CRC32;

Query Match 73.68; Score 64; DB 2; Length 625;
Best Local Similarity 83.3%; Pred. No. 0.0049;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGXLQKNSLLEEVNWL 18
IIIIIIIIIIIIIIIIIIII
DB 28 AGCLQKNSLLEEVNWL 45

RESULT 3
O71306 PRELIMINARY; PRT; 836 AA.
AC O71306;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DNA POLYMERASE.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retroviral viruses; Hepadnaviridae; Avihepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ALBERTA;
RA FISCHER K.P., STICKNEY J., TIPPLES G.A., TYRRELL D.L.J.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047045; AAC06354.1; -
DR PFAM; PF00336; DNA_pol_viral_C; 1.
DR PFAM; PF00242; DNA_pol_viral_N; 1.
DR PFAM; PF00078; rvt; 1.
SQ SEQUENCE 836 AA; 95217 MW; DE377A65 CRC32;

Query Match 52.9%; Score 46; DB 12; Length 836;
Best Local Similarity 61.1%; Pred. No. 8.3;
Matches 11; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 AGXLQKNSLLEEVNWL 16
:IIIIIIIIIIIIIIIIII
DB 409 SGASDKNSPLEENVWYL 426

RESULT 4
Q03765 PRELIMINARY; PRT; 786 AA.
AC Q03765;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DNA-DIRECTED DNA POLYMERASE (EC 2.7.7.7).
GN POL.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retroviral viruses; Hepadnaviridae; Avihepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RA THOMAS R.F., NEWBOLD J.E.;
RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; M06677; AAA62819.1; -
DR PFAM; PF00336; DNA_pol_viral_C; 1.
DR PFAM; PF00242; DNA_pol_viral_N; 1.
DR PFAM; PF00078; rvt; 1.
KW DNA-directed DNA polymerase; DNA replication.
SQ SEQUENCE 786 AA; 89752 MW; AC31E045 CRC32;

Query Match 52.9%; Score 46; DB 12; Length 786;
Best Local Similarity 61.1%; Pred. No. 7.8;
Matches 11; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 AGXLQKNSLLEEVNWL 16

DB 359 SGASDKNSPLEENVWYL 376
:IIIIIIIIIIIIIIIIII

RESULT 5
Q66400 PRELIMINARY; PRT; 836 AA.
AC Q66400;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE POLYMERASE.
OS Duck hepatitis B virus (DHBV).
OC Viruses; Retroviral viruses; Hepadnaviridae; Avihepadnavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INDIAN;
RA MONSHI A., PANDA S.K.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; X74623; CAA52700.1; -
DR PFAM; PF00336; DNA_pol_viral_C; 1.
DR PFAM; PF00242; DNA_pol_viral_N; 1.
DR PFAM; PF00078; rvt; 1.
SQ SEQUENCE 836 AA; 95496 MW; 48ECBE66 CRC32;

Query Match 52.9%; Score 46; DB 12; Length 836;
Best Local Similarity 61.1%; Pred. No. 8.3;
Matches 11; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 AGXLQKNSLLEEVNWL 16
:IIIIIIIIIIIIIIIIII
DB 409 SGASDKNSPLEENVWYL 426

RESULT 6
O45151 PRELIMINARY; PRT; 558 AA.
AC O45151;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE C17B7.8 PROTEIN.
GN C17B7.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FAVELLO A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans*;
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WILSON R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF047655; AAC04402.1; -
 DR PFAM: PF01431; Peptidase_M13; 1.
 SQ SEQUENCE 558 AA; 64066 MW; F57ED4B2 CRC32;

Query Match 51.7%; Score 45; DB 5; Length 558;
 Best Local Similarity 53.3%; Pred. No. 8.1;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 4 LOKNSLLEEVWYLAL 18
 | | | : | | : | |
 Db 211 LMKNNTEIEIWSSAL 225

RESULT 7
 Q921M3 PRELIMINARY; PRT; 413 AA.
 AC Q921M3;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE INTERFERON-G INDUCED GTPASE.
 GN IIGP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RA BOEHM U., GUETHLEIN L., KLAMP T., OZBEK K., SCHAUB A., FUETTERER A.,
 RA PFEFFER K., HOWARD J.C.;
 RT "Two families of GTPases dominate the complex cellular response to
 RT interferon-g.";
 RL J. Immunol. 161:6715-6723(1998).
 DR EMBL: AJ007971; CA07798.1; -
 SQ SEQUENCE 413 AA; 47571 MW; 9D2DC770 CRC32;

Query Match 49.4%; Score 43; DB 11; Length 413;
 Best Local Similarity 61.5%; Pred. No. 13;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 4 LOKNSLLEEVWYL 16
 | | | | | : | | | |
 Db 378 LPRNSFLKEIFYL 390

RESULT 8
 ID P91515 PRELIMINARY; PRT; 494 AA.
 AC P91515;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE COSMID T28A11.
 GN T28A11.17.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSICOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA ROHLFING T.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U80027; AAC48126.1; -
 DR PFAM: PF01431; Peptidase_M13; 1.
 SQ SEQUENCE 494 AA; 56390 MW; EFB5E9F4 CRC32;

Query Match 48.3%; Score 42; DB 5; Length 494;
 Best Local Similarity 46.7%; Pred. No. 24;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 4 LOKNSLLEEVWYLAL 18
 | | | : | | : | |
 Db 168 LMKNNTEIEIWSSAL 182

RESULT 9
 O97283 PRELIMINARY; PRT; 3085 AA.
 AC O97283;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE MAL3P7.10 PROTEIN.
 GN MAL3P7.10.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RA OLIVER K., BOWMAN S., HARRIS D., LAWSON D., QUAIL M., BARRELL B.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL034559; CAB39029.1; -
 SQ SEQUENCE 3085 AA; 360515 MW; 39DBFF93 CRC32;

Query Match 48.3%; Score 42; DB 5; Length 3085;
 Best Local Similarity 60.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 6 KNSLLEEVWY 15
 | | | : | |
 Db 1578 KNSLISNIWY 1587

RESULT 10
 O97299 PRELIMINARY; PRT; 1542 AA.
 AC O97299;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE MAL3P7.37 PROTEIN.
 GN MAL3P7.37.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RA OLIVER K., BOWMAN S., HARRIS D., LAWSON D., QUAIL M., BARRELL B.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL034559; CAB39046.1; -
 SQ SEQUENCE 1542 AA; 184460 MW; 242A4DE9 CRC32;

Query Match 48.3%; Score 42; DB 5; Length 1542;

Best Local Similarity 47.1%; Pred. No. 75;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 GXLQNSLLEEVWYL 18
| | | | | | | | | |
DB 136 GGFQNNKLQKQWLLLL 152

RESULT 11
P74225 PRELIMINARY; PRT; 179 AA.
AC P74225;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE HYPOTHETICAL 20.2 KD PROTEIN.
GN CYP GENE.
OS Chironomus thummi (Midge).
OC Eukaryota; Metazoa; Arthropoda; Insecta;
OC Pterygota; Diptera; Nematocera; Chironomidae;
OC Chironominae; Chironomus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90913; BAA18319.1; -;
KW Hypothetical protein.
SQ SEQUENCE 179 AA; 20180 MW; 1DFC5D1C CRC32;

Query Match 47.1%; Score 41; DB 2; Length 179;
Best Local Similarity 61.5%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 LQKNSLLEEVWYL 16
| | | | | | | | | |
DB 50 LASNGLLEAOWYL 62

RESULT 12
O62524 PRELIMINARY; PRT; 3512 AA.
AC O62524;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CYP GENE.
GN CYP.
OS Chironomus thummi (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Diptera; Nematocera; Chironomidae; Chironominae;
OC Chironominae; Chironomus.
RN [1]
RP SEQUENCE FROM N.A.
RC KRAEMER C., SCHMIDT E.R.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82317; CAA57760.1; -;
DR PFAM; PF00400; WD40.1;
SQ SEQUENCE 3512 AA; 390083 MW; 7C281ABF CRC32;

Query Match 47.1%; Score 41; DB 5; Length 3512;

Best Local Similarity 46.7%; Pred. No. 2.6e+02;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 LQKNSLLEEVWYL 18
| | | | | | | | | |
DB 1044 LQNSVFEFFYIV 1058

RESULT 13
P72877 PRELIMINARY; PRT; 756 AA.
AC P72877;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE EXOPOLYSACCHARIDE EXPORT PROTEIN.
GN EPSB.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90901; BAA16893.1; -;
SQ SEQUENCE 756 AA; 83640 MW; D1E949ED CRC32;

Query Match 46.0%; Score 40; DB 2; Length 756;
Best Local Similarity 72.7%; Pred. No. 80;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 LQKNSLLEEVW 14
| | | | | | | | | |
DB 112 LQSNILLEPWW 122

RESULT 14
O94520 PRELIMINARY; PRT; 193 AA.
AC O94520;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE HYPOTHETICAL 21.6 KD PROTEIN.
GN SPCCL281.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA VOLCKAERT G., WOOD V., RAJANDREAM M.A., BARRELL B.G.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035218; CAA22824.1; -;
KW Hypothetical protein.
SQ SEQUENCE 193 AA; 21551 MW; 5F11C119 CRC32;

Query Match 46.0%; Score 40; DB 3; Length 193;
Best Local Similarity 57.1%; Pred. No. 20;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 LQNSLLEEVWYLA 17
| | | : | | : | | |
Db 66 LQDLLLLKAWELA 79

RESULT 15
O01437 PRELIMINARY; PRT; 252 AA.
AC O01437;
DT 01-JUL-1997 (TEMBLrel. 04, Created)
DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE SIMILAR TO DROSOPHILA RLC1 GENE PRODUCT.
GN B0261.4
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATRELLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA SCHEET P., GATTUNG S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U97016; AAB52351.1;
SQ SEQUENCE 252 AA; 29775 MW; 81A2F108 CRC32;

Query Match 46.0%; Score 40; DB 5; Length 252;
Best Local Similarity 46.7%; Pred. No. 26;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 LQNSLLEEVWYLA 18
| : | | : | | : |
Db 95 LKNSDLKLYVYCL 109

Search completed: November 13, 1999, 12:55:39
Job time: 3038 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 19:00:00 ; Search time 104.22 Seconds
(without alignments)
4.545 Million cell updates/sec

Title: US-08-913-430-13
Perfect score: 105
Sequence: 1 AKNDFAPSIQGYKKIAHEL 20

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq.36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	105	100.0	20	1 W01025	Mycoplasma 72-75 k
2	96	91.4	627	1 W62451	Mycoplasma hyopneu
3	44	41.9	894	1 W86810	WpA protein. New
4	42	40.0	1087	1 R96028	P. gingivalis haem
5	42	40.0	1358	1 R96032	P. gingivalis hagD
6	42	40.0	1358	1 W69494	Haemagglutinin pro
7	42	40.0	1087	1 W69486	Haemagglutinin pro
8	41	39.0	254	1 P94684	Amino acid sequenc
9	41	39.0	382	1 W23969	Connexin protein C
10	39.5	37.6	774	1 W29323	DNA polymerase wit
11	39.5	37.6	1829	1 W29322	DNA polymerase wit
12	39	37.1	2391	1 R55694	Carbamoyl-phosphat
13	39	37.1	1418	1 W15264	Bloom's syndrome act
14	39	37.1	1417	1 W31548	Bloom's syndrome B
15	39	37.1	739	1 W31549	Bloom's syndrome B
16	39	37.1	1417	1 W31550	Bloom's syndrome B
17	39	37.1	1417	1 W31551	Bloom's syndrome B
18	39	37.1	1004	1 W52830	A tumour suppresso
19	39	37.1	1004	1 W37856	Human polyhomeotic
20	39	37.1	323	1 W68497	Treponima pallidum
21	38.5	36.7	377	1 W05411	Human H74 protein.
22	38.5	36.7	495	1 W31629	Aspergillus oryzae
23	38	36.2	1280	1 R04868	Protein encoded by
24	38	36.2	1280	1 P70452	Sequence encoded b
25	38	36.2	1280	1 R42997	Sequence encoded b
26	38	36.2	1280	1 R63624	Human p-glycoprote
27	38	36.2	1280	1 W44073	Human multidrug re
28	38	36.2	1280	1 W48997	Wild-type human p
29	38	36.2	1280	1 W48998	N-terminal single
30	38	36.2	1280	1 W48999	C-terminal single
31	38	36.2	1280	1 W49000	Human P glycoprote
32	38	36.2	1279	1 W80294	Mutated human P-gl
33	38	36.2	1974	1 W98391	H. pylori GHPO 57
34	37	35.2	165	1 R07360	Rat melanin-concen
35	37	35.2	812	1 R10047	abaA gene of Asper
36	37	35.2	970	1 R42451	Enzyme involved in
37	37	35.2	970	1 R99460	Biosynthetic enzym
38	37	35.2	474	1 W00669	Glutathione synthe
39	37	35.2	799	1 W19212	Corn starch branch
40	37	35.2	970	1 W37048	S. putrefaciens EP
41	37	35.2	799	1 W56489	zea mays starch br
42	37	35.2	711	1 W53103	Streptococcus pneu
43	36.5	34.8	775	1 R24813	Sequence of Pfu py

ALIGNMENTS

RESULT 1
W01025
ID W01025 standard; Peptide; 20 AA.
DT 19-JAN-1997 (first entry)
DE Mycoplasma 72-75 kDa protective antigen peptide CNBR Fl.
KW Antigen; vaccine; mycoplasmal pneumonia; swine enzootic pneumonia;
diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
PN W09628472-A1.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (UYME) UNIV MELBOURNE.
PI Doughty SW, Lee R, Walker J;
DR WPI: 96-433763/43
PT putative protective antigens against Mycoplasma - used for the
PT detection, prevention or treatment of Mycoplasma infections, esp. M.
PT hyopneumoniae in swine
PS Claim 8; Page 27; 43pp; English.
CC A 72-75 kDa putative protective antigen against Mycoplasma
CC contains the N-terminal sequence given in W01024 and the internal
CC CNBR fragments given in W01025-27. The antigen was isolated from
CC Mycoplasma hyopneumoniae cells using antibody probes enriched with
CC Mycoplasma-specific antibodies. Other protective antigens were
CC also identified (see also W01028-37). Protective antigens and
CC antibodies can be used in vaccines for preventing or treating
CC mycoplasmal infections, partic. M. hyopneumoniae infections in
CC swine. They can also be used for diagnosis.
SQ Sequence 20 AA;

Query Match 100.0%; Score 105; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.4e-11; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

QY 1 AKNDFAPSIQGYKKIAHEL 20
|||||
DB 1 AKNDFAPSIQGYKKIAHEL 20

RESULT 2
W62451
ID W62451 standard; Protein; 627 AA.
DT 01-OCT-1998 (first entry)
DE Mycoplasma hyopneumoniae P65 surface antigen.
KW Mycoplasma hyopneumoniae; surface lipoprotein P65; structural gene;
antigen; vaccine; swine; enzootic pneumoniae; diagnosis; pig;
immune response; mycoplasmal pneumonia.
OS Mycoplasma hyopneumoniae.
PN US5788962-A.
PD 04-AUG-1998.
PF 28-AUG-1996; 703947.
PR 17-JAN-1995; US-373957.
PR 28-AUG-1996; US-703947.
PA (UMOR) UNIV MISSOURI.
PI McIntosh WA, Wise KS;
DR WPI: 98-446004/38.
PR N-PSDB; V39862.
PT Vaccine against mycoplasmal pneumonia in pigs - containing
PT Mycoplasma hyopneumoniae P65 surface antigen fusion protein
PS Claim 1; Fig 2; 29pp; English.
CC A vaccine has been developed for protecting pigs against mycoplasmal
CC pneumonia caused by Mycoplasma hyopneumoniae. The vaccine contains an
CC immunogenic fusion protein comprising a first amino acid sequence fused

CC to a second amino acid sequence, where the first sequence is the present
 CC sequence of 627 amino acids, which is a P65 surface antigen. The
 CC sequence given in V39862 represents the Mycoplasma hyopneumoniae
 CC surface lipoprotein P65 structural gene, which encodes the P65 surface
 CC antigen. The vaccine is used to induce an immune response in pigs
 CC against mycoplasma pneumonia caused by Mycoplasma hyopneumoniae.
 SQ Sequence 627 AA;

Query Match 91.4%; Score 96; DB 1; Length 627;
 Best Local Similarity 90.9%; Pred. NO. 2.2e-08;
 Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKNFDAPSIOGYKKIAHEL 20
 ||||| ||||| ||||| :
 Db 318 AKNFDHPSIOGYKKIAHQL 337

RESULT 3

W86810 ID W86810 standard; Protein; 894 AA.
 AC W86810;
 DT 08-MAR-1999 (first entry)
 DE WpA protein.
 KW Wall protease A; CWBP23; CWBP52; native; heterologous; recombinant;
 KW wall-associated protease; wprA.
 OS Bacillus subtilis.
 PN W09849328-A1.
 PD 05-NOV-1998.
 PF 24-APR-1998; G01051.
 PR 26-APR-1997; GB-008452.
 PA (NOVO) NOVO-NORDISK AS.
 PA (UYNE-) UNIV NEWCASTLE-UPON-TYNE.
 PI Hansen C, Harwood CR, Jorgensen S, Kristensen T,
 PI Stevenson K;
 DR WPI; 99-024065/02.
 DR N-PSDB; V71268.

PT New Bacillus strain microorganisms lacking the wall protease A gene
 PT - are useful for increased heterologous or recombinant protein
 PT production and secretion
 PS Disclosure; Fig 1; 50pp; English.
 CC The present sequence is the wall protease A (WpA) protein, derived from
 CC Bacillus subtilis. The gene encoding this protein produces a 96 kDa
 CC polypeptide which upon synthesis and export gives rise to two cell
 CC wall-bound proteins, CWBP23 and CWBP52. In the method of the invention
 CC this gene has been altered by deletion, insertion, mutation and/or
 CC substitution so that production and/or function of the wprA gene product
 CC is affected to facilitate production of a native, heterologous or
 CC recombinant polypeptide. Absence of a wall-associated protease such as
 CC wprA, leads to an increase in secreted protein production. Thus this new
 CC bacterial strain can be used for the production of native, heterologous
 CC and/or recombinant polypeptides.
 SQ Sequence 894 AA;

Query Match 41.9%; Score 44; DB 1; Length 894;
 Best Local Similarity 38.9%; Pred. NO. 21;
 Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 NFDAPSIOGYKKIAHEL 20
 | : | : | : | : | : | :
 Db 400 NVEFAEPVOEYKSLANDI 417

RESULT 4

R96028 ID R96028 standard; Protein; 1087 AA.
 AC R96028;
 DT 04-SEP-1996 (first entry)
 DE P. gingivalis haemagglutinin hagD.
 KW Haemagglutinin; hagD; periodontal disease; vaccine; antibody.
 OS Porphyromonas gingivalis strain FDC381.
 PN W09617936-A2.

PD 13-JUN-1996.
 PR 11-DEC-1995; U16108.
 PR 09-DEC-1994; US-353485.
 PA (UABR-) UAB RES FOUND.
 PA (UYFL) UNIV FLORIDA.
 PI Han N, Lantz M, Lepine G, Patti JM, Progulske-Fox A;
 PI Tunwasorn S;
 DR WPI; 96-287181/29.
 DR N-PSDB; T30652.
 PT Porphyromonas gingivalis genes and proteins - used in the detection
 PT and vaccination against periodontal disease
 PS Claim 5; Page 65-68; 153pp; English.
 CC P. gingivalis 381 haemagglutinin hagD (R96028) was identified as
 CC the product of a gene (T30652) isolated from a P. gingivalis 318
 CC genomic library. The haemagglutinin (see also R96032) can be obtd.
 CC from transformed host cells and used as a vaccine to protect humans
 CC or animals against periodontal disease. Expression in Salmonella
 CC cells allows prodn. of a live vaccine. The haemagglutinin can also
 CC be used to detect the presence of anti-P. gingivalis antibodies and
 CC to raise monoclonal antibodies for diagnostic appln.
 SQ Sequence 1087 AA;

Query Match 40.0%; Score 42; DB 1; Length 1087;
 Best Local Similarity 38.9%; Pred. NO. 58;
 Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDAPSIOGYKKIAHEL 20
 || | : | : | : | : | :
 Db 859 NFLITPKVEGAKKITKYV 876

RESULT 5

R96032 ID R96032 standard; Protein; 1358 AA.
 AC R96032;
 DT 04-SEP-1996 (first entry)
 DE P. gingivalis hagD haemagglutinin.
 KW Haemagglutinin; hagD; periodontal disease; vaccine; antibody.
 OS Porphyromonas gingivalis strain FDC381.
 PN W09617936-A2.
 PD 13-JUN-1996.
 PR 11-DEC-1995; U16108.
 PR 09-DEC-1994; US-353485.
 PA (UABR-) UAB RES FOUND.
 PA (UYFL) UNIV FLORIDA.
 PI Han N, Lantz M, Lepine G, Patti JM, Progulske-Fox A;
 PI Tunwasorn S;
 DR WPI; 96-287181/29.
 DR N-PSDB; T30655.

PT Porphyromonas gingivalis genes and proteins - used in the detection
 PT and vaccination against periodontal disease
 PS Claim 5; Page 125-129; 153pp; English.
 CC P. gingivalis 381 haemagglutinin hagD (R96032) was identified as
 CC the product of the second open reading frame of the hagD gene
 CC (T30655) derived from P. gingivalis 318 genomic DNA. A first
 CC open reading frame coded for hagD protease (see also R96031).
 CC The protease and haemagglutinin can be obtd. from transformed host
 CC cells and used in vaccines to protect humans or animals against
 CC periodontal disease. Expression in Salmonella cells allows prodn.
 CC of live vaccines. The haemagglutinin and protease can also be used
 CC to detect the presence of anti-P. gingivalis antibodies and to raise
 CC monoclonal antibodies for diagnostic appln.
 SQ Sequence 1358 AA;

Query Match 40.0%; Score 42; DB 1; Length 1358;
 Best Local Similarity 38.9%; Pred. NO. 75;
 Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDAPSIOGYKKIAHEL 20
 || | : | : | : | : | :
 Db 1130 NFLITPKVEGAKKITKYV 1147

```

CC vaccines against periodontal disease.
SQ Sequence 1087 AA;

Query Match 40.0%; Score 42; DB 1; Length 1087;
Best Local Similarity 38.9%; Pred. NO. 58;
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDFAFSIQYKKIAHEL 20
   || :||| :|:
Db 859 NFLITPKVEGAKKITVKV 876

RESULT 8
P94684
ID P94684 standard; protein; 254 AA.
AC P94684;
DT 29-JUN-1990 (first entry)
DE Amino acid sequence encoded by part of human mannose binding protein
KW (hMBP) genomic DNA
OS Homo sapiens.
PN W08901519-A.
PD 23-AUG-1987; U02591.
PF 05-AUG-1988; U02591.
PR 20-AUG-1987; US-087628.
PA (CHIL) Childrens Hospital.
PI Ezekowitz RAB;
DR WPI: 89-068874/09.
DR N-PSDB; N94351.
PT New DNA encoding human mannose binding protein or fragment -
PT for treating infections, used for deriving diagnostic antibodies
PS Fig 3, part 2; 34pp; English.
CC A 650 bp C-terminal PstI fragment of an hMBP cDNA clone was used as
CC probe for human genomic library. Clones which hybridised under
CC stringent conditions were isolated. The positively identified clones
CC were plaque purified and their nucleic acid sequence determined
CC (which is given in n91080 and n94351) and the corresponding amino acid
CC sequence given in 3 reading frames. In p94684, X represents where
CC there was a stop codon on the corresponding DNA. The patent claims
CC an engineered nucleic acid which encodes 20 or more contiguous
CC amino acids of hMBP. The engineered nucleic acid encodes a peptide
CC with more than 75% homology to a fragment of 30 or more amino acids
CC of hMBP and contains 90 or more bases hybridisable with hMBP. The
CC patent claims (expression) vectors and cells contg. it, peptides
CC encoded by it and antibodies (Ab) against these peptides. Peptides
CC encoded by it, opt. fused to a toxin, are used to treat animals
CC infected with bacterial, fungal or viral (including HIV)
CC pathogens. The Abs are useful in diagnosing susceptibility to infection
CC (by ELISA).
SQ Sequence 254 AA;

Query Match 39.0%; Score 41; DB 1; Length 254;
Best Local Similarity 72.7%; Pred. NO. 17;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AKNEDFAPSIQ 11
   ||| |||||:|
Db 94 AVNHFAPSLQ 104

RESULT 9
W23969
ID W23969 standard; Protein; 382 AA.
AC W23969;
DT 06-JUL-1998 (first entry)
DE Connexin protein Cx43.
KW Connexin Cx43; gap junction protein; cardiac conduction;
KW sick sinus syndrome; heart block; bradycardia; tachycardia;
KW brady-tachy syndrome; therapy.
OS Homo sapiens.
PN W09802150-A1.

CC vaccines against periodontal disease.
SQ Sequence 1087 AA;

Query Match 40.0%; Score 42; DB 1; Length 1358;
Best Local Similarity 38.9%; Pred. NO. 75;
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDFAFSIQYKKIAHEL 20
   || :||| :|:
Db 1130 NFLITPKVEGAKKITVKV 1147

RESULT 7
W69486
ID W69486 standard; Protein; 1087 AA.
AC W69486;
DT 22-DEC-1998 (first entry)
DE Haemagglutinin protein hagd.
KW Haemagglutinin protein; periodontal disease; vaccine; hagd.
OS Porphyromonas gingivalis.
PN US5824791-A.
PD 20-OCT-1998.
PF 11-DEC-1995; 570311.
PR 11-DEC-1995; US-570311.
PR 08-SEP-1988; US-241640.
PR 25-JAN-1991; US-647119.
PR 09-DEC-1994; US-353485.
PA (UABR-) UAB RES FOUND.
PA (UYFL) UNIV FLORIDA.
PI Han N, Lantz M, Lepine G, Patti JM, Progulske-Fox A,
PI Tumwasorn S;
DR WPI: 98-582627/49.
DR N-PSDB; V58880.
PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin
PT and/or protease poly(peptide(s))
PS Claim 1; Column 145-158; 101pp; English.
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
CC invention. This sequence represents the hagd haemagglutinin protein. The
CC polypeptides are used to produce antibodies to organisms associated with
CC periodontal disease. The antibodies are also used in purification and
CC identification procedures. The genes and polypeptides are used as
CC vaccines against periodontal disease.
SQ Sequence 1358 AA;

Query Match 40.0%; Score 42; DB 1; Length 1358;
Best Local Similarity 38.9%; Pred. NO. 75;
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDFAFSIQYKKIAHEL 20
   || :||| :|:
Db 1130 NFLITPKVEGAKKITVKV 1147

RESULT 7
W69486
ID W69486 standard; Protein; 1087 AA.
AC W69486;
DT 22-DEC-1998 (first entry)
DE Haemagglutinin protein hagd.
KW Haemagglutinin protein; periodontal disease; vaccine; hagd.
OS Porphyromonas gingivalis.
PN US5824791-A.
PD 20-OCT-1998.
PF 11-DEC-1995; 570311.
PR 11-DEC-1995; US-570311.
PR 08-SEP-1988; US-241640.
PR 25-JAN-1991; US-647119.
PR 09-DEC-1994; US-353485.
PA (UABR-) UAB RES FOUND.
PA (UYFL) UNIV FLORIDA.
PI Han N, Lantz M, Lepine G, Patti JM, Progulske-Fox A,
PI Tumwasorn S;
DR WPI: 98-582627/49.
DR N-PSDB; V58873.
PT Isolated Porphyromonas gingivalis genes - encoding haemagglutinin
PT and/or protease poly(peptide(s))
PS Claim 1; Column 57-64; 101pp; English.
CC This sequence is encoded by a Porphyromonas gingivalis gene of the
CC invention. This sequence represents the hagd haemagglutinin protein. The
CC polypeptides are used to produce antibodies to organisms associated with
CC periodontal disease. The antibodies are also used in purification and
CC identification procedures. The genes and polypeptides are used as
```

PD 22-JAN-1998.
 PF 04-APR-1997; U06103.
 PR 17-JUL-1996; US-682277.
 PA (MEDT) MEDTRONIC INC.
 PI Morissette J, Stokes KB;
 DR WPI; 98-110323/10.
 DR N-PSDB; V04682.
 PT System for delivering genetic material to heart - comprises
 PT reservoir, catheter and optionally including mapping electrode,
 PT useful in, e.g. treating conduction disorders such as tachycardia
 PS Claim 16; Page 37-38; 58pp; English.
 CC This polypeptide comprises the gap junction protein connexin Cx43.
 CC The invention provides a delivery system for delivering conduction
 CC protein genetic material (GM) to cardiac cells in localised areas
 CC of the heart in order to modulate cardiac conduction. A system is
 CC provided for delivering connexin proteins (see W23968-70) or
 CC nucleic acid molecules (see V04681-93) encoding connexin proteins
 CC by perfusion or injection to a site in the heart which has been
 CC determined by mapping procedures to have a conduction disturbance.
 CC For cases where conduction is impaired, selected GM is delivered to
 CC cells around the disturbance area, in order to enhance overall
 CC conductivity patterns; in other cases, GM is selected to slow
 CC conduction in affected areas. The method can be used in the
 CC treatment of sick sinus syndrome, heart block, bradycardia,
 CC tachycardia and brady-tachy syndrome. Cx43 is preferably used in
 CC cases of heart block or bradycardia.
 SQ Sequence 382 AA;

Query Match 39.0%; Score 41; DB 1; Length 382;
 Best Local Similarity 50.0%; Pred. No. 27;
 Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

OY 1 AKNDFAPSIQGYKIA--HEL 20
 DB 332 AQPDFDDNQNSKKAAGHEL 353

RESULT 10

W29323
 ID W29323 standard; Protein; 774 AA.
 AC W29323;
 DT 20-APR-1998 (first entry)
 DE DNA polymerase with 3'-5' exonuclease activity.
 KW TY Exon; DSM 10597; thermostable; DNA polymerase;
 KW 3'-5' exonuclease; amplification.
 OS Thermococcus sp.
 PN DE19611759-A1.
 PD 02-OCT-1997.
 PF 25-MAR-1996; 011759.
 PR (BOEF) BOEHRINGER MANNHEIM GMBH.
 PA Antranikian G, Frey B, Niehaus F;
 PI WPI; 97-481494/45.
 DR N-PSDB; T86434.
 DT Thermostable DNA polymerase from Thermococcus sp. TY - useful for
 PT nucleic acid amplification
 PS Claim 1; Pages 9-10; 32pp; German.
 CC The present sequence (TY Exon) is a Thermococcus sp. TY
 CC (DSM 10597) thermostable DNA polymerase with 3'-5' exonuclease
 CC activity.
 CC The enzyme can specifically amplify nucleic acid fragments of up to
 CC 5 kb in high yields, has an activity half-life of 20 minutes at 90
 CC degrees C, has an optimum temperature of 70-80 degrees C, has an
 CC optimum pH of 7.5, exhibits optimum activity at a KCl concentration
 CC of 80-100 mM, is magnesium ion-dependent and is inhibited by
 CC manganese ions.
 SQ Sequence 774 AA;

Query Match 37.6%; Score 39.5; DB 1; Length 774;
 Best Local Similarity 69.2%; Pred. No. 1.1e+02;
 Matches 9; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

OY 2 KNDFAPSIQGYK 14
 DB 432 KNYDVAP-IVGYK 443

RESULT 11

W29322
 ID W29322 standard; Protein; 1829 AA.
 AC W29322;
 DT 20-APR-1998 (first entry)
 DE DNA polymerase with 3'-5' exonuclease activity.
 KW TYpol Intron; DSM 10597; thermostable; DNA polymerase;
 KW 3'-5' exonuclease; amplification.
 OS Thermococcus sp.
 FH Key Location/Qualifiers
 FT Misc_difference 1116
 FT /label= unknown
 FT /note= "encoded by GGN"
 FT Misc_difference 1118
 FT /label= unknown
 FT /note= "encoded by NTC"
 FT Misc_difference 1123
 FT /label= unknown
 FT /note= "encoded by NTG"
 PN DE19611759-A1.
 PD 02-OCT-1997.
 PF 25-MAR-1996; 011759.
 PR (BOEF) BOEHRINGER MANNHEIM GMBH.
 PA Antranikian G, Frey B, Niehaus F;
 PI WPI; 97-481494/45.
 DR N-PSDB; T86433.
 DT Thermostable DNA polymerase from Thermococcus sp. TY - useful for
 PT nucleic acid amplification
 PS Claim 1; Pages 5-8; 32pp; German.
 CC The present sequence (TYpol Intron) is a Thermococcus sp. TY
 CC (DSM 10597) thermostable DNA polymerase with 3'-5' exonuclease
 CC activity.
 CC The enzyme can specifically amplify nucleic acid fragments of up to
 CC 5 kb in high yields, has an activity half-life of 20 minutes at 90
 CC degrees C, has an optimum temperature of 70-80 degrees C, has an
 CC optimum pH of 7.5, exhibits optimum activity at a KCl concentration
 CC of 80-100 mM, is magnesium ion-dependent and is inhibited by
 CC manganese ions.
 SQ Sequence 1829 AA;

Query Match 37.6%; Score 39.5; DB 1; Length 1829;
 Best Local Similarity 69.2%; Pred. No. 2.8e+02;
 Matches 9; Conservative -1; Mismatches 2; Indels 1; Gaps 1;

OY 2 KNDFAPSIQGYK 14
 DB 792 KNYDVAP-IVGYK 803

RESULT 12

R55694
 ID R55694 standard; Protein; 2391 AA.
 AC R55694;
 DT 06-DEC-1994 (first entry)
 DE Carbamoyl-phosphate-synthetase II.
 KW Carbamoyl-phosphate-synthetase II; CPSII; pscPSII gene;
 KW malaria.
 OS Plasmodium falciparum.
 FH Key Location/Qualifiers
 FT domain 1..690
 FT /note= "glutamine-amidotransferase domain"
 FT domain 1..270
 FT /note= "structural subdomain"
 FT peptide 271..482
 FT /note= "insert sequence"

```

FT domain 483..690
FT /note= "glutaminase subdomain"
FT 691..2391
FT domain
FT /note= "carbamoyl-phosphate-synthase domain"
FT 691..1254
FT domain
FT /note= "ATP binding subdomain CPSa"
FT 1255..1857
FT peptide /note= "insert sequence"
FT 1858..2391
FT domain
FT /note= "ATP binding subdomain CPSb"
FT W09412643-A.
PN 09-JUN-1994.
PD 02-DEC-1993; AU0617.
PF 03-DEC-1992; AU-006206.
PR 16-DEC-1992; AU-006380.
PI (UNIX ) UNISEARCH LTD.
PI Flores MV, Osullivan WJ, Stewart TS;
DR WPI: 94-200271/24.
DR N-PSDB; Q62924.
PT Nucleic acid encoding carbamoyl phosphate synthetase II -
PT isolated from Plasmodium falciparum, used to develop prods. for
PT the treatment of malaria.
PS Disclosure: Page 6-16; 31pp; English.
CC The cDNA sequence encoding the carbamoyl-phosphate-transferase II
CC (CPSII) of Plasmodium falciparum was determined. The cDNA encodes
CC a protein that includes 2 insert sequences not found in other CPSII
CC proteins. The first separates the putative structural subdomain and
CC the glutaminase subdomain of the glutamine-amidotransferase subunit
CC of CPSII, while the second separates 2 ATP binding subdomains of the
CC CPSII subunit. CPSa and CPSb.
CC Sequence 2391 AA;

Query Match 37.1%; Score 39; DB 1; Length 2391;
Best Local Similarity 58.3%; Pred. No. 4.5e+02;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 9 STQGYKKIAHEL 20
DB 891 SLQGWKEIEYEL 902

RESULT 13
W15264
ID W15264 standard; Protein: 1418 AA.
AC W15264;
DE 26-JAN-1998 (first entry)
DE Bloom syndrome active BLM protein.
KW BLM; Bloom syndrome; BS; mutant; probe; PCR primer; cancer;
KW Therapy; diagnosis; vector.
OS Homo sapiens.
PN W09717979-Al.
PD 22-MAY-1997.
PF 15-NOV-1996; U19046.
PR 15-NOV-1995; US-559303.
PA (NYBL-) NEW YORK BLOOD CENT INC.
PI Ellis N, German J, Groden J;
DR WPI: 97-289051/26.
DR N-PSDB; T67013.
PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT genes - for gene therapy with nucleic acid encoding active BLM
PT protein to treat Bloom's syndrome and cancer in general
PS Claim 58; Fig 2; 51pp; English.
CC This active BLM protein is encoded by a 4437 bp BLM gene sequence.
CC This is used in diagnosing Bloom's syndrome where 2 mutated BLM genes
CC or the absence of a wild-type BLM gene in the nucleic acid of a subject
CC binding a 1.3 cm region on chromosome 15. cDNA from this region was
CC isolated, a 847 bp fragment selected and used to screen a HeLa cDNA
CC library. The longest clone H1 was isolated and extended by PCR to isolate
CC this BLM gene sequence. Delivery of a functional BLM gene to bone marrow
CC cells, e.g. via the viral vectors, is used to treat or prevent the onset
CC of Bloom's syndrome. Identification of the BLM gene and its products

CC should assist in the development of therapeutic and diagnostic agents for
CC cancer.
CC Sequence 1418 AA;

Query Match 37.1%; Score 39; DB 1; Length 1418;
Best Local Similarity 31.0%; Pred. No. 2.5e+02;
Matches 9; Conservative 5; Mismatches 5; Indels 10; Gaps 1;

QY 1 AKNFDFAPSIOGY-----KKIAHE 19
DB 614 AQNINFSESIQNYTKSAQNLSARNLKHE 642

RESULT 14
W31548
ID W31548 standard; Protein: 1417 AA.
AC W31548;
DE 27-JAN-1998 (first entry)
DE Bloom's syndrome BLM mutated protein.
KW BLM; Bloom's syndrome; BS; mutant; therapy; diagnosis.
OS Homo sapiens.
PH Key Location/Qualifiers
FT Misc-difference 672
FT /label= Q672R
FT /note= "wild.type Gln is replaced by Arg"
FT W09717979-Al.
PN 22-MAY-1997.
PF 15-NOV-1996; U19046.
PR 15-NOV-1995; US-559303.
PA (NYBL-) NEW YORK BLOOD CENT INC.
PI Ellis N, German J, Groden J;
DR WPI: 97-289051/26.
DR N-PSDB; T93392.
PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT genes - for gene therapy with nucleic acid encoding active BLM
PT protein to treat Bloom's syndrome and cancer in general
PS Disclosure: Page -; 51pp; English.
CC This is a mutated BLM protein encoded by a mutant gene sequence
CC isolated from an American/European Bloom's Syndrome sufferer designated
CC "139(VIKre)". The substitution of the bp A at the position 2089 of the
CC wild type H1-5' gene to the bp G results in the amino acid Arginine at
CC the position 672.
CC The substitution of the base A at the position 888 of the wild type H1-5'
CC gene to the base T results in a stop codon at amino acid position 272.
CC This was one of the seven unique mutations which were identified in
CC a study of 10 people with Bloom's Syndrome. Based on the various
CC mutations, diagnostic tests for Bloom's Syndrome have been developed
CC which use standard sequence analysis techniques to detect the presence
CC of 2 mutated BLM genes or the absence of a wild-type BLM gene.
CC Note: The present sequence does not appear in the specification; it
CC has been made by modifying the H1-5' wild type BLM sequence which is
CC provided in Figure 2 (W15264).
CC Sequence 1417 AA;

Query Match 37.1%; Score 39; DB 1; Length 1417;
Best Local Similarity 31.0%; Pred. No. 2.5e+02;
Matches 9; Conservative 5; Mismatches 5; Indels 10; Gaps 1;

QY 1 AKNFDFAPSIOGY-----KKIAHE 19
DB 614 AQNINFSESIQNYTKSAQNLSARNLKHE 642

RESULT 15
W31549
ID W31549 standard; Protein: 739 AA.
AC W31549;
DE 27-JAN-1998 (first entry)
DE Bloom's syndrome BLM mutated protein.
KW BLM; Bloom's syndrome; BS; mutant; truncation; therapy; diagnosis.
OS Homo sapiens.

```

FH	Key
FT	Location/Qualifiers
FT	Protein
FT	1..739
FT	/note= "Deletion of ATCTGA from the wild type H1-5'
FT	BLM gene and insertion of TAGATTC results in
FT	the insertion of the novel codons for LDSR after
FT	amino acid position 735 followed by a stop codon"
FT	"
FN	WO9717979-A1.
PD	22-MAY-1997.
PD	U19046.
PF	15-NOV-1996; US-559303.
PR	15-NOV-1995; US-559303.
PA	(NYBL-) NEW YORK BLOOD CENT INC.
PI	Ellis N, German J, Groden J;
PI	WPI: 97-289051/26.
DR	N-PDSB; T93393.
DR	D
PT	Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM
PT	genes - for gene therapy with nucleic acid encoding active BLM
PT	protein to treat Bloom's syndrome and cancer in general
PS	Disclosure; Page -: 51pp; English.
PS	This is a truncated BLM protein encoded by a mutant gene sequence
CC	isolated from a Ashkenazi Jewish Bloom's Syndrome sufferer designated
CC	'15(Maro)". The deletion of ATCTGA from the wild type H1-5' BLM gene and
CC	insertion of TAGATTC results in the insertion of the novel codons for
CC	LDSR after amino acid position 735 followed by a stop codon.
CC	This was one of the seven unique mutations which were identified in
CC	a study of 10 people with Bloom's Syndrome. Based on the various
CC	mutations, diagnostic tests for Bloom's Syndrome have been developed
CC	which use standard sequence analysis techniques to detect the presence
CC	of 2 mutated BLM genes or the absence of a wild-type BLM gene.
CC	Note: The present sequence does not appear in the specification; it
CC	has been made by modifying the H1-5' wild type BLM sequence which is
CC	provided in Figure 2 (W15264).
SQ	Sequence 739 AA;

Search completed: November 13, 1999, 19:00:02
Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:35 ; Search time 64.87 Seconds
(without alignments)
3.524 Million cell updates/sec

Title: US-08-913-430-13

Perfect score: 105

Sequence: 1 AKNDFAPSIQGYKKIAHEL 20

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/PCTUS9_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	96	91.4	627	US-08-703-947-2	Sequence 2, Appl
2	43	41.0	3200	US-08-477-451-8	Sequence 8, Appl
3	42	40.0	1087	US-08-570-311-8	Sequence 8, Appl
4	42	40.0	1358	US-08-570-311-27	Sequence 27, Appl
5	42	40.0	1087	US-08-353-485-8	Sequence 8, Appl
6	39	37.1	1417	US-08-559-303B-78	Sequence 78, Appl
7	39	37.1	2391	US-08-446-855A-2	Sequence 2, Appl
8	38	36.2	1279	US-08-784-649A-2	Sequence 2, Appl
9	38	36.2	1280	US-08-583-276-19	Sequence 19, Appl
10	37	35.2	970	US-08-375-709-7	Sequence 7, Appl
11	37	35.2	970	US-08-752-929-7	Sequence 6, Appl
12	37	35.2	205	US-08-829-110-6	Sequence 6, Appl
13	37	35.2	205	US-08-748-483-5	Sequence 5, Appl
14	36.5	34.8	775	US-07-966-278-1	Sequence 1, Appl
15	36.5	34.8	775	US-08-424-921-1	Sequence 1, Appl
16	36.5	34.8	1022	US-08-271-364A-8	Sequence 1, Appl
17	36.5	34.8	776	US-08-688-649-37	Sequence 37, Appl
18	36.5	34.8	1022	US-08-222-715B-27	Sequence 27, Appl
19	36.5	34.8	533	US-08-325-488-2	Sequence 2, Appl
20	36.5	34.8	775	US-08-556-355A-1	Sequence 1, Appl
21	36.5	34.8	409	US-08-743-130A-2	Sequence 2, Appl
22	36.5	34.8	409	US-08-743-130A-39	Sequence 39, Appl
23	36.5	34.8	775	US-07-803-627A-1	Sequence 1, Appl
24	36	34.3	440	US-08-483-140-27	Sequence 27, Appl
25	36	34.3	258	US-08-449-045C-2	Sequence 2, Appl
26	36	34.3	1104	US-08-327-832-5	Sequence 5, Appl
27	36	34.3	440	US-08-485-938A-31	Sequence 31, Appl
28	36	34.3	258	US-08-435-605A-2	Sequence 2, Appl
29	36	34.3	194	US-08-435-605A-6	Sequence 6, Appl
30	36	34.3	192	US-08-435-605A-8	Sequence 8, Appl
31	36	34.3	240	US-08-435-605A-13	Sequence 13, Appl
32	36	34.3	194	US-08-435-605A-14	Sequence 14, Appl
33	36	34.3	240	US-08-435-605A-15	Sequence 15, Appl
34	36	34.3	194	US-08-435-605A-16	Sequence 16, Appl
35	36	34.3	240	US-08-435-605A-49	Sequence 49, Appl
36	36	34.3	240	US-08-435-605A-50	Sequence 50, Appl
37	36	34.3	240	US-08-435-605A-51	Sequence 51, Appl
38	36	34.3	240	US-08-435-605A-52	Sequence 52, Appl
39	36	34.3	240	US-08-435-605A-53	Sequence 53, Appl

Sequence 54, Appl
Sequence 55, Appl
Sequence 56, Appl
Sequence 57, Appl
Sequence 4, Appl
Sequence 5, Appl

40 36 34.3 194 2 US-08-435-605A-54
41 36 34.3 194 2 US-08-435-605A-55
42 36 34.3 240 2 US-08-435-605A-56
43 36 34.3 179 2 US-08-435-605A-57
44 36 34.3 3729 2 US-08-804-227C-4
45 36 34.3 1104 2 US-08-828-584-5

ALIGNMENTS

RESULT 1
US-08-703-947-2
; Sequence 2, Application US/08703947
; Patent No. 5788962
; GENERAL INFORMATION:
; APPLICANT: Wise, Kim S.
; APPLICANT: McIntosh, Mark A.
; TITLE OF INVENTION: DNA Sequences Coding for Mycoplasma
; TITLE OF INVENTION: Hypopneumoniae Surface Antigens, in
; TITLE OF INVENTION: Corresponding Proteins and Use in
; TITLE OF INVENTION: Vaccines and Diagnostic Procedures
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace J. Fishel
; STREET: 929 Fee Fee Road, Suite 100
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63043
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
; COMPUTER: Hewlett-Packard Vectra
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,947
; FILING DATE: 28-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/373,957
; FILING DATE: January 17, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fishel, Grace J.
; REGISTRATION NUMBER: 25864
; REFERENCE/DOCKET NUMBER: UVM 8141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 878-0440
; TELEFAX: (314) 275-7693
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 amino acid residues
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: predicted amino acid sequence of complete
; DESCRIPTION: 627 residues of the p65 lipoprotein, derived
; DESCRIPTION: from the nucleic acid sequence
; HYPOTHETICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: whole polypeptide
; ORIGINAL SOURCE:
; ORGANISM: Mycoplasma hypopneumoniae
; STRAIN: J
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE: unicellular bacterium
; CELL LINE:
; ORGANELLE:

IMMEDIATE SOURCE:
LIBRARY: Genomic in Charon 4A, GEM12
CLONE: MhpJ25, MhpJ35, MhpJG35, p2J25.1, p2J25.14,
CLONE: p2JG35.1, p2JG35.12, p2JG35.13, p2JG35.14
POSITION IN GENOME:
CHROMOSOME/SEGMENT: single chromosome
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: 627 amino acid sequence representing
NAME/KEY: complete sequence (including signal
NAME/KEY: sequence) of surface lipoprotein P65
LOCATION: entire derived coded sequence
IDENTIFICATION METHOD: clone identified by immunodetection of
IDENTIFICATION METHOD: protein product with antiserum specific for
IDENTIFICATION METHOD: P65; residue sequence deduced from nucleic
IDENTIFICATION METHOD: acid sequence
OTHER INFORMATION: immunogenic surface lipoprotein of no known
OTHER INFORMATION: function; C-terminus exposed on external
OTHER INFORMATION: surface of cell; N-terminal signal sequence
OTHER INFORMATION: (first 29 amino acid residues) cleaved during
OTHER INFORMATION: lipid modification process
PUBLICATION INFORMATION:
AUTHORS: Mary F. Kim, Manijeh B. Heidari, Susan J.
TITLES: Identification and Mapping of an
TITLES: Immunogenic Region of Mycoplasma
TITLES: hypnemoniae p65 Surface Lipoprotein
TITLES: Expressed in Escherichia coli from a Cloned
TITLES: Genomic Fragment
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 8
PAGES: 2637-2643
DATE: August 1990
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: From 1 to 627

US-08-703-947-2

Query Match 91.4%; Score 96; DB 2; Length 627;
Best Local Similarity 90.0%; Pred. No. 5.3e-09;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKNDFAPSIQGYKKIAHEL 20
Db 318 AKNDFHPSIQGYKKIAHQL 337

RESULT 2
US-08-477-451-8
; Sequence 8, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/477,451

; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3200 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-451-8

Query Match 41.0%; Score 43; DB 2; Length 3200;
Best Local Similarity 58.8%; Pred. No. 72;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 FDFAPSIQGYKKIAHEL 20
Db 1513 FCFLPFFQGLIKHAHEL 1529

RESULT 3
US-08-570-311-8
; Sequence 8, Application US/08570311
; Patent No. 5824791
; GENERAL INFORMATION:
; APPLICANT: Progulske-Fox, Ann
; APPLICANT: Tumwasorn, Somying
; APPLICANT: Lepine, Guylaine
; APPLICANT: Han, Naiming
; APPLICANT: Lantz, Marilyn
; APPLICANT: Patti, Joseph
; TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ted W. Whitlock
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,311
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,485
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/647,119
; FILING DATE: 25-JAN-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/241,640
; FILING DATE: 08-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF15.C3

TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1087 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-8

Query Match 40.0%; Score 42; DB 2; Length 1087;
Best Local Similarity 38.9%; Pred. No. 30;
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDFAFSIOGYKKIAHEL 20
|| |::| |||:::
Db 859 NFLITPRVEGAKKITKYK 876

RESULT 4
US-08-570-311-27
Sequence 27, Application US/08570311
Patent No. 5824791

GENERAL INFORMATION:
APPLICANT: Proguiske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,311
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1358 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-570-311-27

Query Match 40.0%; Score 42; DB 2; Length 1358;
Best Local Similarity 38.9%; Pred. No. 39;
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDFAFSIOGYKKIAHEL 20
|| |::| |||:::
Db 1130 NFLITPRVEGAKKITKYK 1147

RESULT 5
US-08-353-485-8
Sequence 8, Application US/08353485
Patent No. 5830710

GENERAL INFORMATION:
APPLICANT: Proguiske-Fox, Ann
APPLICANT: Tumwasorn, Somying
APPLICANT: Lepine, Guylaine
APPLICANT: Han, Naiming
APPLICANT: Lantz, Marilyn
APPLICANT: Patti, Joseph
TITLE OF INVENTION: Cloned Porphyromonas gingivalis Genes
and Probes for the Detection of Periodontal Disease
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ted W. Whitlock
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,485
FILING DATE: 09-DEC-1994
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/647,119
FILING DATE: 25-JAN-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/241,640
FILING DATE: 08-SEP-1988
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF15.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1087 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-353-485-8

Query Match 40.0%; Score 42; DB 2; Length 1087;
Best Local Similarity 38.9%; Pred. No. 30;
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDFAFSIOGYKKIAHEL 20
|| |::| |||:::

Db 859 NFLITPKVGAKKITYK 876

RESULT 6

US-08-559-303B-78
; Sequence 78, Application US/08559303B
; Patent No. 5824501
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
; APPLICANT: GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,303B
FILING DATE: NOVEMBER 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 63475/65
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:
LENGTH: 1417
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: OTHER NUCLEIC ACID
HYPOTHETICAL: YES
ANTI-SENSE: NO
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:

US-08-559-303B-78

Query Match 37.1%; Score 39; DB 2; Length 1417;
Best Local Similarity 31.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 5; Mismatches 5; Indels 10; Gaps 1;

QY 1 AKNFDAPSIQGY-----KKIAHE 19

Db 614 AQINFSESIQNYTDKSAQNLASRLKHE 642

RESULT 7

US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 NO. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-855A-2

Query Match 37.1%; Score 39; DB 2; Length 2391;
Best Local Similarity 58.3%; Pred. No. 2.6e+02;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 9 SIQYKKIAHEL 20

Db 891 SLOGWREIEYEL 902

RESULT 8

US-08-784-649A-2
; Sequence 2, Application US/08784649A
; Patent No. 5830697
; GENERAL INFORMATION:
; APPLICANT: Sikic, Branimir I
; APPLICANT: Chen, Gang
; TITLE OF INVENTION: P-GLYCOPROTEIN MUTANT RESISTANT TO
; TITLE OF INVENTION: CYCLOSPORIN MODULATION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/784,649A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: Reg. No. 5830697 36,677
REFERENCE/DOCKET NUMBER: 06037/007001

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-322-5070
TELEFAX: 415-854-0875
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1279 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-784-649A-2

Query Match 36.2%; Score 38; DB 2; Length 1279;
Best Local Similarity 44.4%; Pred. No. 1.9e+02;
Matches 8; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 2 KNDFAPSIQGYKKIAHE 19

Db 366 KIIDNKPSIDSYSGHK 383

RESULT 9

US-08-583-276-19
Sequence 19, Application US/08583276
Patent No. 5837536

GENERAL INFORMATION:

APPLICANT: McDonagh, Kevin T.
APPLICANT: Nienhuis, Arthur
APPLICANT: Tolstoshev, Paul
TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
TITLE OF INVENTION: SELECTION OF CELLS TRANSDUCED WITH SUCH GENES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
ADDRESSEE: Cecchi & Stewart
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: DM4 V2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/583,276
FILING DATE: 05-JAN-1996
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/332,444
FILING DATE: 31-OCT-1994
APPLICATION NUMBER: 07/887,712
FILING DATE: 22-MAY-1992

INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1280 amino acids
TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE:

DESCRIPTION: protein

US-08-583-276-19

Query Match 36.2%; Score 38; DB 2; Length 1280;
Best Local Similarity 44.4%; Pred. No. 1.9e+02;
Matches 8; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 2 KNDFAPSIQGYKKIAHE 19

| | | | | | | |

Db 367 KIIDNKPSIDSYSGHK 384

RESULT 10

US-08-375-709-7
Sequence 7, Application US/08375709
Patent No. 5683898

GENERAL INFORMATION:

APPLICANT: YAZAWA, Kazunaga
APPLICANT: YAMADA, Akiko
APPLICANT: KATO, Seishi
APPLICANT: KONDO, Kiyosi
TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid
TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of
TITLE OF INVENTION: Eicosapentaenoic Acid
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/375,709
FILING DATE: 20-JAN-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/178,251
FILING DATE: 14-MAY-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-147945

FILING DATE: 15-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: WEGNER, Harold C.

REGISTRATION NUMBER: 25,258

REFERENCE/DOCKET NUMBER: 53466/150/AAOK

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 970 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-375-709-7

Query Match 35.2%; Score 37; DB 1; Length 970;
Best Local Similarity 63.6%; Pred. No. 2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 NFDFAPSIQGY 13

| | | | | | | |

Db 339 NFDFTDNIQFY 349

RESULT 11

US-08-752-929-7
Sequence 7, Application US/08752929
Patent No. 5798259

GENERAL INFORMATION:

APPLICANT: YAZAWA, Kazunaga
APPLICANT: YAMADA, Akiko
APPLICANT: KATO, Seishi
APPLICANT: KONDO, Kiyosi

;; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing
;; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic
;; TITLE OF INVENTION: Acid
;; NUMBER OF SEQUENCES: 25
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: 3000 K Street, N.W., Suite 500
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20007-5109
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/752,929
;; FILING DATE: 20-NOV-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/375,709
;; FILING DATE: 20-JAN-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/178,251
;; FILING DATE: 14-MAY-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 4-147945
;; FILING DATE: 15-MAY-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: WEGNER, Harold C.
;; REGISTRATION NUMBER: 25,258
;; REFERENCE/DOCKET NUMBER: 53466/150/AAOK
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 970 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-752-929-7

Query Match 35.2%; Score 37; DB 2; Length 970;
Best Local Similarity 63.6%; Pred. No. 2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 NFDFAPIQGY 13
Db 339 NFDFTDNIQFY 349

RESULT 12
US-08-829-110-6
; Sequence 6, Application US/08829110
; Patent No. 5882890
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL REGULATORS OF G-PROTEIN
; TITLE OF INVENTION: SIGNALING
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/829,110
;; FILING DATE: Filed Herewith
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Billings, Lucy J.
;; REGISTRATION NUMBER: 36,749
;; REFERENCE/DOCKET NUMBER: PF-0259 US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-855-0555
;; TELEFAX: 415-845-4166
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 205 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; LIBRARY: GenBank
;; CLONE: 1216373
US-08-829-110-6

Query Match 35.2%; Score 37; DB 2; Length 205;
Best Local Similarity 53.3%; Pred. No. 33;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 KNDFAPSIQGYKKI 16
Db 87 ENIDFWISCEYKKI 101

RESULT 13
US-08-748-483-5
; Sequence 5, Application US/08748483
; Patent No. 595314
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL REGULATOR OF CELL SIGNALING
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,483
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0157 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1216373
US-08-748-483-5

Query Match 35.2%; Score 37; DB 2; Length 205;
Best Local Similarity 53.3%; Pred. No. 33;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 KNDFAPSIQGYK 16
DB 87 ENDFWISCEYK 101

RESULT 14

US-07-966-278-1
Sequence 1, Application US/07966278
Patent No. 5489523
GENERAL INFORMATION:
APPLICANT: Mathur, Eric A
TITLE OF INVENTION: EXONUCLEASE-DEFICIENT THERMOSTABLE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Thomas Fitting
STREET: 12526 High Bluff Road, Suite 300
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92130

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/966,278
FILING DATE: 19921226
CLASSIFICATION: 435

PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: STG0133P
TELEPHONE: 619-792-3680
TELEFAX: 619-792-8477

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 775 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-966-278-1

Query Match 34.8%; Score 36.5; DB 1; Length 775;
Best Local Similarity 53.8%; Pred. No. 1.9e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 KNDFAPSIQGYK 14
DB 430 KNYDIAPOV-GHK 441

RESULT 15

US-08-424-921-1
Sequence 1, Application US/08424921
Patent No. 5545552
GENERAL INFORMATION:
APPLICANT: Mathur, Eric A
TITLE OF INVENTION: PURIFIED THERMOSTABLE PYROCOCUS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Bingham & Fitting
STREET: 12526 High Bluff Road, Suite 300
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92130

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,921
FILING DATE: 19-APR-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,627
FILING DATE: 02-DEC-1991
APPLICATION NUMBER: US 07/620,568
FILING DATE: 03-DEC-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,073
FILING DATE: 19-FEB-1991
APPLICATION NUMBER: US 07/776,552
FILING DATE: 15-OCT-1991

ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: STG0100P
TELEPHONE: 619-792-3680
TELEFAX: 619-792-8477

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 775 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-424-921-1

Query Match 34.8%; Score 36.5; DB 1; Length 775;
Best Local Similarity 53.8%; Pred. No. 1.9e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 KNDFAPSIQGYK 14
DB 430 KNYDIAPOV-GHK 441

Search completed: November 13, 1999, 10:56:35
Job time: 1364 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:40 ; Search time 251.81 Seconds
(without alignments)
5.028 Million cell updates/sec

Title: US-08-913-430-13

Perfect score: 105

Sequence: 1 AKNFDAPSIOGYKKIAHEL 20

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA:*

1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	20	13	US-08-913-430-13
2	96	91.4	627	6	US-08-373-957-2
3	43	41.0	3200	12	US-08-851-486-8
4	42	40.0	154	6	US-08-388-137A-10
5	42	40.0	149	19	US-60-136-242-156
6	42	40.0	149	19	US-60-140-803-529
7	42	40.0	149	19	US-60-142-845-621
8	42	40.0	296	19	US-60-143-134-179
9	41	39.0	382	1	PCT-US97-06103-4
10	41	39.0	382	10	US-08-682-277A-4
11	41	39.0	211	16	US-09-248-796-15399
12	41	39.0	211	19	US-60-096-409-15399
13	41	39.0	116	23	US-09-417-507-27580
14	40	38.1	789	15	US-09-107-532-6734
15	40	38.1	425	15	US-09-134-000-3861
16	40	38.1	769	15	US-09-134-000-5119
17	40	38.1	292	17	US-09-328-352-7538
18	39	37.1	2391	7	US-08-446-855-2
19	39	37.1	1417	9	US-08-559-303A-78

20	39	37.1	1417	9	US-08-580-539-78	Sequence 78, Appl
21	39	37.1	1417	9	US-08-594-242-78	Sequence 78, Appl
22	39	37.1	311	10	US-08-625-811-1075	Sequence 1075, Ap
23	39	37.1	1417	10	US-08-632-175-78	Sequence 78, Appl
24	39	37.1	1417	11	US-08-781-891-78	Sequence 78, Appl
25	39	37.1	1004	13	US-08-916-352-2	Sequence 2, Appli
26	39	37.1	1004	13	US-08-916-377-2	Sequence 2, Appli
27	39	37.1	5588	14	US-09-036-987A-6	Sequence 6, Appli
28	39	37.1	702	15	US-09-107-532-6866	Sequence 6866, Ap
29	39	37.1	765	16	US-09-270-767-43804	Sequence 43804, A
30	39	37.1	5588	17	US-09-370-700-6	Sequence 2, Appli
31	39	37.1	502	19	US-60-031-569-2	Sequence 2, Appli
32	38.5	36.7	377	10	US-08-630-315-36	Sequence 36, Appl
33	38.5	36.7	377	10	US-08-630-915A-36	Sequence 35, Appl
34	38.5	36.7	114	13	US-08-943-667-35	Sequence 4, Appli
35	38.5	36.7	495	14	US-09-079-415-4	Sequence 4699, Ap
36	38.5	36.7	313	19	US-60-128-476-4699	Sequence 382, App
37	38	36.2	1974	1	PCT-US98-06371-382	Sequence 1148, Ap
38	38	36.2	84	10	US-08-625-811-1148	Sequence 2, Appli
39	38	36.2	1280	11	US-08-752-447-2	Sequence 2, Appli
40	38	36.2	1279	11	US-08-784-649-2	Sequence 308, App
41	38	36.2	529	12	US-08-816-454-308	Sequence 3138, Ap
42	38	36.2	457	12	US-08-827-356-3138	Sequence 5620, Ap
43	38	36.2	465	15	US-09-107-532-5620	Sequence 3669, Ap
44	38	36.2	241	15	US-09-134-000-3669	Sequence 5488, Ap
45	38	36.2	333	15	US-09-134-000-5488	

ALIGNMENTS

RESULT 1
US-08-913-430-13
; Sequence 13, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 13
; TYPE: PRT
; LENGTH: 20
; ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-13

Query Match 100.0%; Score 105; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 1e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKNFDAPSIOGYKKIAHEL 20
|||||
DB 1 AKNFDAPSIOGYKKIAHEL 20

RESULT 2
US-08-373-957-2
; Sequence 2, Application US/08373957
; GENERAL INFORMATION:
; APPLICANT: Wise, Kim S.
; APPLICANT: McIntosh, Mark A.
; TITLE OF INVENTION: DNA Sequences Coding for Mycoplasma
; TITLE OF INVENTION: Hyopneumoniae Surface Antigens,
; TITLE OF INVENTION: Corresponding Proteins and Use in Vaccines

TITLE OF INVENTION: and Diagnostic Procedures
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grace J. Fishel
STREET: 929 Fee Fee Road, Suite 100
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63043
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
COMPUTER: Hewlett-Packard Vectra
OPERATING SYSTEM: MS-DOS version 3.3
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/373,957
FILING DATE: January 17, 1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fishel, Grace J.
REGISTRATION NUMBER: 25864
REFERENCE/DOCKET NUMBER: UVM 8141
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 878-0440
TELEFAX: (314) 275-7693
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 627 amino acid residues
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: predicted amino acid sequence of complete
DESCRIPTION: 627 residues of the p65 lipoprotein, derived
DESCRIPTION: from the nucleic acid sequence
HYPOTHETICAL: no
ANTI-SENSE: no whole polypeptide
FRAGMENT TYPE: no whole polypeptide
ORIGINAL SOURCE:
ORGANISM: Mycoplasma hyopneumoniae
STRAIN: J
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE: unicellular bacterium
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY: Genomic in Charon 4A, GEM12
CLONE: MhpJ25, MhpJ35, MhpJG35, p2J25,
CLONE: p2J25.1, p2J25.14, p2JG35.1, p2JG35.12,
CLONE: p2JG35.13, p2JG35.14
POSITION IN GENOME:
CHROMOSOME/SEGMENT: single chromosome
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: 627 amino acid sequence representing
NAME/KEY: complete sequence (including signal
NAME/KEY: sequence) of surface lipoprotein p65
LOCATION: entire derived coded sequence
IDENTIFICATION METHOD: clone identified by immunodetection of
IDENTIFICATION METHOD: protein product with antiserum specific for
IDENTIFICATION METHOD: p65; residue sequence deduced from nucleic
IDENTIFICATION METHOD: acid sequence
OTHER INFORMATION: immunogenic surface lipoprotein of no known
OTHER INFORMATION: function; C-terminus exposed on external
OTHER INFORMATION: surface of cell; N-terminal signal sequence

OTHER INFORMATION: (first 29 amino acid residues) cleaved during
OTHER INFORMATION: lipid modification process
PUBLICATION INFORMATION:
AUTHORS: Mary F. Kim, Manijeh B. Heidari, Susan J.
AUTHORS: Stull, Mark A. McIntosh, and Kim S. Wise
TITLE: Identification and Mapping of an
TITLE: Immunogenic Region of Mycoplasma
TITLE: hyopneumoniae p65 Surface Lipoprotein
TITLE: Expressed in Escherichia coli from a Cloned
TITLE: Genomic Fragment
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 8
PAGES: 2637-2643
DATE: August 1990
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: From 1 to 627
US-08-373-957-2

Query Match 91.4%; Score 96; DB 6; Length 627;
Best Local Similarity 90.0%; Pred. No. 2.7e-08;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKNDFAPSIQGYKTAHEL 20
Db 318 AKNDFPSPSIQGYKTAHQL 337

RESULT 3
US-08-851-486-8
Sequence 8, Application US/08851486
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
TITLE OF INVENTION: Helicobacter Pylori CagI Region
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,486
FILING DATE: 05-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/425,194
FILING DATE: 20-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0335.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2708
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3200 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-851-486-8

```

; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000016
; CURRENT APPLICATION NUMBER: US/60/136,242
; CURRENT FILING DATE: 1999-05-26
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 156
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Drosophila
US-60-136-242-156

Query Match 40.0%; Score 42; DB 19; Length 149;
Best Local Similarity 37.5%; Pred. No. 15;
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 FDFAPSIQGYKKIAHE 19
| | | | | : : |
DB 28 FQFVPFVWGFAQLCHE 43

RESULT 6
US-60-140-803-529
; Sequence 529, Application US/60140803
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000042
; CURRENT APPLICATION NUMBER: US/60/140,803
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 529
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Drosophila
US-60-140-803-529

Query Match 40.0%; Score 42; DB 19; Length 149;
Best Local Similarity 37.5%; Pred. No. 15;
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 FDFAPSIQGYKKIAHE 19
| | | | | : : |
DB 28 FQFVPFVWGFAQLCHE 43

RESULT 7
US-60-142-845-621
; Sequence 621, Application US/60142845
; GENERAL INFORMATION:
; APPLICANT: Kerlavage, Anthony
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000048
; CURRENT APPLICATION NUMBER: US/60/142,845
; CURRENT FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 704
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 621
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(149)
; OTHER INFORMATION: Xaa = Any Amino Acid

```

US-60-142-845-621

Query Match 40.0% Score 42; DB 19; Length 149;
Best Local Similarity 37.5%; Pred. No. 15;
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 FDFAPSIQYKKIAHE 19
| | | | | : : : : :
Db 28 FQVFFVWGFAQLCHE 43

RESULT 8

US-60-145-134-179
; Sequence 179, Application US/60145134
; GENERAL INFORMATION:
; APPLICANT: Bonazzoli, Vivien
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; FILE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES THERE
; CURRENT APPLICATION NUMBER: CLO00058
; CURRENT FILING DATE: 1999-07-22
; NUMBER OF SEQ ID NOS: 342
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 179
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Drosophila
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(296)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-60-145-134-179

Query Match 40.0% Score 42; DB 19; Length 296;
Best Local Similarity 37.5%; Pred. No. 34;
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 FDFAPSIQYKKIAHE 19
| | | | | : : : : :
Db 193 FQVFFVWGFAQLCHE 208

RESULT 9

PCT-US97-06103-4
; Sequence 4, Application PC/TUS9706103
; GENERAL INFORMATION:
; APPLICANT: Stokes, Kenneth
; APPLICANT: Morissette, Jos e
; TITLE OF INVENTION: SYSTEM FOR GENETICALLY TREATING CARDIAC
; TITLE OF INVENTION: CONDUCTION DISTURBANCES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/06103
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul K. Legaard
; REGISTRATION NUMBER: 38,534
; REFERENCE/DOCKET NUMBER: MEDT-0059

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 382 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
PCT-US97-06103-4

Query Match 39.0% Score 41; DB 1; Length 382;
Best Local Similarity 50.0%; Pred. No. 70;
Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 AKNFDAPSIQYKKIA--HEL 20
| : | | | | | : | | | | |
Db 332 AQPFDFFDDNQNSKLAAGHEL 353

RESULT 10

US-08-682-277A-4
; Sequence 4, Application US/08682277A
; GENERAL INFORMATION:
; APPLICANT: Stokes, Kenneth
; APPLICANT: Morissette, Jos e
; TITLE OF INVENTION: SYSTEM AND METHOD FOR GENETICALLY TREATING
; TITLE OF INVENTION: CARDIAC CONDUCTION DISTURBANCES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and Norris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,277A
; FILING DATE: Herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul K. Legaard
; REGISTRATION NUMBER: 38,534
; REFERENCE/DOCKET NUMBER: MEDT-0028/P-3569
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 382 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
US-08-682-277A-4

Query Match 39.0% Score 41; DB 10; Length 382;
Best Local Similarity 50.0%; Pred. No. 70;
Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 AKNFDAPSIQYKKIA--HEL 20
| : | | | | | : | | | | |
Db 332 AQPFDFFDDNQNSKLAAGHEL 353

RESULT 11
US-09-248-796-15399

```
; Sequence 15399, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 15399
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (199),(205),(210)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796-15399

Query Match          39.0%; Score 41; DB 16; Length 211;
Best Local Similarity 70.0%; Pred. No. 34;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      6 FAPSIQYKK 15
      ||| | |||
Db     144 FAPDINGYKR 153

RESULT 12
US-60-096-409-15399
; Sequence 15399, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 15399
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (199),(205),(210)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-60-096-409-15399

Query Match          39.0%; Score 41; DB 19; Length 211;
Best Local Similarity 70.0%; Pred. No. 34;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      6 FAPSIQYKK 15
      ||| | |||
Db     144 FAPDINGYKR 153

RESULT 13
US-09-417-507-27580
; Sequence 27580, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 27580
; LENGTH: 116
```

```
; TYPE: PRT
; ORGANISM: A. fumigatus
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (110),(43),(45),(60),(97)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-417-507-27580

Query Match          39.0%; Score 41; DB 23; Length 116;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY      1 AKNFDAPSIOGYKKIAHEL 20
      | |||| | : | | |
Db     27 ALNFDFACLIFSHELIIXHL 46

RESULT 14
US-09-107-532-6734
; Sequence 6734, Application US/09107532
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR
; NUMBER OF SEQUENCES: 7308
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085598
; FILING DATE: May 14, 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6734:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 789 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...789
US-09-107-532-6734

Query Match          38.1%; Score 40; DB 15; Length 789;
Best Local Similarity 63.6%; Pred. No. 2.6e+02;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

QY 2 KNFDFAPSIQG 12
|:|:|:|:|
Db 769 KSFEFAPANQG 779

RESULT 15
US-09-134-000-3861
; Sequence 3861, Application US/09134000A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS
; FILE REFERENCE: GTC-005
; CURRENT APPLICATION NUMBER: US/09/134,000A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 6810
; SEQ ID NO 3861
; TYPE: PRT
; LENGTH: 425
; ORGANISM: Enterococcus faecalis
US-09-134-000-3861

Query Match 38.1%; Score 40; DB 15; Length 425;
Best Local Similarity 61.5%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 FAPSIQGYKKIAH 18
|:|:|:|:|
Db 385 FGPSFGQKGI AH 397

Search completed: November 13, 1999, 05:08:42
Job time: 10868 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:08:07 ; Search time 75.45 Seconds
(without alignments)
10.620 Million cell updates/sec

Title: US-08-913-430-13

Perfect score: 105

Sequence: 1 AKNFDAPSIQGYKKIAHEL 20

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	43.8	381	1 A48171	gap junction prote
2	45	42.9	328	2 S24344	glucose transport
3	45	42.9	682	2 B70121	hypothetical prote
4	44	41.9	894	2 F69730	cell wall-associat
5	44	41.9	168	2 H64237	hypothetical prote
6	44	41.9	211	2 E70476	2-acetylglucosaminidase
7	43.5	41.4	567	2 A42509	E6R protein - vacc
8	43.5	41.4	567	2 E35928	hypothetical prote
9	43.5	41.4	567	2 I36841	E6R protein - vari
10	43	41.0	382	1 A39802	gap junction prote
11	43	41.0	162	2 S26176	transcription elon
12	42	40.0	369	1 A23559	dihydroorotate ox
13	42	40.0	382	1 S00532	gap junction prote
14	42	40.0	212	2 D69864	hypothetical prote
15	42	40.0	392	2 D64433	hypothetical prote
16	42	40.0	436	2 S64744	hypothetical prote
17	42	40.0	201	2 D71190	hypothetical prote
18	41	39.0	382	1 A35853	gap junction prote
19	41	39.0	240	2 S63218	hypothetical prote
20	40	38.1	379	1 A34575	gap junction prote
21	40	38.1	616	1 VX2RNA	beta-1,2-glucan ex
22	40	38.1	613	2 JC6033	mosquitocidal prot
23	40	38.1	865	2 B69074	formate dehydrogen
24	40	38.1	998	2 A36368	transcription fact
25	40	38.1	358	2 H71088	hypothetical prote
26	39.5	37.6	505	2 FC5129	trans-cinnamate 4-
27	39.5	37.6	396	2 PC4236	trans-cinnamate 4-
28	39.5	37.6	484	2 S73490	glutamate--tRNA li
29	39.5	37.6	144	2 G65170	heat shock protein
30	39	37.1	685	1 EF8Y52	suppressor 2 prote
31	39	37.1	1417	1 A47570	Bloom's syndrome r
32	39	37.1	317	2 B41671	iron transport pro
33	39	37.1	197	2 S41182	phage-related repl
34	39	37.1	308	2 H71904	hypothetical prote
35	39	37.1	308	2 C64608	yabC protein homol
36	39	37.1	934	2 E64235	hypothetical prote
37	39	37.1	162	2 D65140	hypothetical prote
38	39	37.1	507	2 T01844	probable sugar tra
39	39	37.1	948	2 A57640	retinoblastoma pro

RAE-28 - mouse
membrane antigen t
klotho protein - R
trans-cinnamate 4-
hypothetical prote
conserved hypothet

ALIGNMENTS

```
RESULT 1
A48171
gap junction protein Cx43 - chicken
N:Alternate names: connexin 43; gap junction alpha-1 protein
C:Species: Gallus gallus (chicken)
C>Date: 03-Feb-1994 #sequence_revision 02-Jun-1994 #text_change 05-Sep-1997
C:Accession: A48171
R:Musil, L.S.; Beyer, E.C.; Goodenough, D.A.
J. Membr. Biol. 116, 163-175, 1990
A:Title: Expression of the gap junction protein connexin43 in embryonic chick lens: m
A:Reference number: A48171; NUID:90339470
A:Accession: A48171
A:Molecule type: mRNA
A:Residues: 1-381 <MUS>
A:Cross-references: GB:M29003; NID:g211642; PID:g211643
C:Superfamily: gap junction protein
C:Keywords: gap junction; phosphoprotein; transmembrane protein
F:1-18/Domain: intracellular #status predicted <INT1>
F:19-41/Domain: transmembrane #status predicted <TM1>
F:42-76/Domain: extracellular #status predicted <EE1>
F:77-98/Domain: transmembrane #status predicted <TM2>
F:99-153/Domain: intracellular #status predicted <INT2>
F:154-184/Domain: transmembrane #status predicted <TM3>
F:185-206/Domain: extracellular #status predicted <EE2>
F:207-229/Domain: transmembrane #status predicted <TM4>
F:230-381/Domain: intracellular #status predicted <INT3>
F:296,364,367,368,371,372/Binding site: phosphate (Ser) (covalent) (by protein kinase

Query Match 43.8%; Score 46; DB 1; Length 381;
Best Local Similarity 54.5%; Pred. No. 3.7;
Matches 12; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 1 AKNFDAPSIQGYKKIA--HEL 20
   |:||||| | |||||
Db 331 AQFFDFADEHQNTKLASGHEL 352

RESULT 2
S24344
glucose transport protein Glut7 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 04-Sep-1998
C:Accession: S24344
R:Waddell, I.D.; Zomerschoe, A.G.; Voice, M.W.; Burchell, A.
Biochem. J. 286, 173-177, 1992
A:Title: Cloning and expression of a hepatic microsomal glucose transport protein. Co
A:Reference number: S24344; NUID:92392274
A:Accession: S24344
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-528 <WAD>
A:Cross-references: EMBL:X66031; NID:g56266; PID:g56267
C:Superfamily: glucose transport protein
C:Keywords: transmembrane protein

Query Match 42.9%; Score 45; DB 2; Length 528;
Best Local Similarity 58.3%; Pred. No. 7.9;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 9 SIQGYKKIAHEL 20
```

||||| :|||
Db 260 SLOCYDEVSHL 271

RESULT 3
B70121
hypothetical protein BB0170 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
C:Accession: B70121
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943
A:Accession: B70121
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-682 <KLE>
A:Cross-references: GB:AE001128; GB:AE000783; NID:g2688057; PID:g2688070; TIGR:BB0170
A:Experimental source: strain B31

Query Match 42.9%; Score 45; DB 2; Length 682;
Best Local Similarity 61.5%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 KNDFAPSIQGYK 14
||| ||| |||
Db 317 KNYFKPSKNGYK 329

RESULT 4
F69730
cell wall-associated protein precursor wprA - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
C:Accession: F69730
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton
C.; Bron, S.; Brouillet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galled
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: F69730
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-894 <KUN>
A:Cross-references: GB:299109; GB:AL009126; NID:g2633260; PID:el183079; PID:g2633413
A:Experimental source: strain 168
C:Genetics:
A:Gene: wprA
C:Superfamily: subtilisin homology
F:453-664/Domain: subtilisin homology <SBT>

Query Match 41.9%; Score 44; DB 2; Length 894;
Best Local Similarity 38.9%; Pred. No. 21;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 NFDFAPSIQGYKTAHEL 20

||||| :||| :||| :|||
Db 400 NVEFAEPVQYKSLANDI 417

RESULT 5
H64237
hypothetical protein MG342 - Mycoplasma genitalium (SGC3)
C:Species: Mycoplasma genitalium
C>Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 10-Oct-1997
C:Accession: H64237
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346
A:Accession: H64237
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-168 <TIGR>
A:Cross-references: GB:U39718; GB:L43967; NID:g1046047; PID:g1046048; TIGR:MG342
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3

Query Match 41.9%; Score 44; DB 2; Length 168;
Best Local Similarity 53.3%; Pred. No. 3.2;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 KNDFAPSIQGYKKI 16
||||| :||| :|||
Db 153 ENDFKPFIEQVKKL 167

RESULT 6
E70476
2-acylglycerophosphoethanolamine acyltransferase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
C:Accession: E70476
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196866
A:Accession: E70476
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-211 <AQF>
A:Cross-references: GB:AE000770; NID:g2984274; PID:g2984277; GB:AE000657
A:Experimental source: strain VFS
C:Genetics:
A:Gene: aas

Query Match 41.9%; Score 44; DB 2; Length 211;
Best Local Similarity 46.2%; Pred. No. 4.2;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 8 PSIQGYKTAHEL 20
||||| :||| :||| :|||
Db 185 PSVKYKRVANEI 197

RESULT 7
A42509
E6R protein - vaccinia virus (strain Copenhagen)
C:Species: vaccinia virus
A:Note: host Homo sapiens (man)
C>Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 08-Apr-1994
C:Accession: A42509
R:Johnson, G.P.

submitted to GenBank, June 1990
A:Reference number: A33172
A:Accession: A42509
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-567 <JOH>

Query Match 41.4%; Score 43.5; DB 2; Length 567;
Best Local Similarity 45.0%; Pred. No. 15;
Matches 9; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY 2 KNF-DFAPSIQGYKKIAHEL 20
||: | : || ||: | : ||
Db 94 KNYLSFNSAIQSKYKTVHKL 113

RESULT 8

hypothetical protein E6R - vaccinia virus

C:Species: vaccinia virus
C:Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 09-Sep-1997
A:Accession: E35928
R:Ahn, B.Y.; Gershon, P.D.; Jones, E.V.; Moss, B.
Mol. Cell. Biol. 10, 5433-5441, 1990
A:Title: Identification of rpo30, a vaccinia virus RNA polymerase gene with structural s
A:Reference number: A35928; MUID:90377234
A:Accession: E35928
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-567 <AHN>
A:Cross-references: GB:M63639; NID:g335747; PID:g335753

Query Match 41.4%; Score 43.5; DB 2; Length 567;
Best Local Similarity 45.0%; Pred. No. 15;
Matches 9; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY 2 KNF-DFAPSIQGYKKIAHEL 20
||: | : || ||: | : ||
Db 94 KNYLSFNSAIQSKYKTVHKL 113

RESULT 9

E6R protein - variola virus (strain India-1967)

C:Species: variola virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Sep-1997
A:Accession: I36841

R:Blinov, V.M.
submitted to GenBank, November 1992
A:Description: not shown.
A:Reference number: A36859
A:Accession: I36841
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-567 <BLI>
A:Cross-references: GB:M69198; NID:g456758; PID:g297228

Query Match 41.4%; Score 43.5; DB 2; Length 567;
Best Local Similarity 45.0%; Pred. No. 15;
Matches 9; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY 2 KNF-DFAPSIQGYKKIAHEL 20
||: | : || ||: | : ||
Db 94 KNYLSFNSAIQSKYKTVHKL 113

RESULT 10

A39802

gap junction protein Cx43 - mouse
N:Alternate names: connexin 43; gap junction alpha-1 protein

C:Species: Mus musculus (house mouse)
C:Date: 20-Mar-1992 #sequence_revision 02-Jun-1994 #text_change 18-Sep-1998
C:Accession: A39802; S18110; A49769; S23110
R:Beyer, E.C.; Steinberg, T.H.
J. Biol. Chem. 266, 7971-7974, 1991
A:Title: Evidence that the gap junction protein connexin-43 is the ATP-induced pore o
A:Reference number: A39802; MUID:91217014
A:Accession: A39802
A:Molecule type: mRNA
A:Residues: 1-382 <BEV>
A:Cross-references: GB:M61896; NID:g192678; PID:g192679
R:Ruangsoravatt, C.P.; Morgan, J.L.; Lo, C.W.
submitted to the EMBL Data Library, September 1991
A:Description: Connexin 43 expression in the early mouse embryo: differential distrib
A:Reference number: S18110
A:Accession: S18110
A:Molecule type: mRNA
A:Residues: 1-382 <RUA>
A:Cross-references: EMBL:X61576; NID:g50509; PID:g50510
R:Nishi, M.; Kumar, N.M.; Gilula, N.B.
Dev. Biol. 146, 117-130, 1991
A:Title: Developmental regulation of gap junction gene expression during mouse embryo
A:Reference number: A49769
A:Accession: A49769
A:Molecule type: mRNA
A:Residues: 1-319, 'T', 321-340, 'N', 342-382 <NIS>
A:Cross-references: GB:M63801; NID:g191773; PID:g567197
R:Willecke, K.
submitted to the EMBL Data Library, October 1991
A:Reference number: S23110
A:Accession: S23110
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <WIL>
A:Cross-references: EMBL:X62836; NID:g50522; PID:g50523
C:Superfamily: gap junction protein
C:Keywords: gap junction; phosphoprotein; transmembrane protein
F:19-41/Domain: intracellular #status predicted <INT1>
F:154-184/Domain: transmembrane #status predicted <TM1>
F:42-76/Domain: extracellular #status predicted <EEL>
F:77-98/Domain: transmembrane #status predicted <TM2>
F:99-153/Domain: intracellular #status predicted <INT2>
F:185-206/Domain: extracellular #status predicted <EE2>
F:207-229/Domain: transmembrane #status predicted <TM4>
F:230-382/Domain: intracellular #status predicted <INT3>
F:297,364,365,368,372,373/Binding site: phosphate (Ser) (covalent) (by protein ki

Query Match 41.0%; Score 43; DB 1; Length 382;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;
QY 1 AKNFDFAPSIQGYKKIA--HEL 20
| : ||| | : ||| |||
Db 332 AQPFDPPDSONAKKVAAGHEL 353

RESULT 11

S26176
transcription elongation factor greA - Rickettsia prowazekii

N:Alternate names: RP861
C:Species: Rickettsia prowazekii
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Dec-1998
C:Accession: S26176; E71648; S21472
R:Marks, G.L.; Wood, D.O.
Nucleic Acids Res. 20, 3785, 1992
A:Title: Nucleotide sequence of the Rickettsia prowazekii greA homolog.
A:Reference number: S26176; MUID:92350688
A:Accession: S26176
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-162 <MAR>

A:Cross-references: EMBL:212122; NID:g46345; PID:g46346
 R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.
 A:Reference number: A71630
 A:Accession: E71648
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-162 <AND>
 A:Cross-references: GB:AJ235273; GB:AJ235269; NID:g3861237; PID:el343130; PID:g3861386
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: greA; RP861
 C:Superfamily: transcription elongation factor greB
 C:Keywords: DNA binding; transcription regulation

Query Match 41.0%; Score 43; DB 2; Length 162;
 Best Local Similarity 44.4%; Pred. NO. 4.6;
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDFAFSIQYKKIAHEL 20

Db 2 NTFPPITAGFKKLEHEL 19

RESULT 12

A23559

dihydroorotate oxidase (EC 1.3.3.1) - slime mold (*Dictyostellium discoideum*)

N:Alternate names: dihydroorotate dehydrogenase

C:Species: *Dictyostellium discoideum*

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993

C:Accession: A23559

R:Jacquet, M.; Kalekine, M.; Boy-Marcotte, E.

Biochimie 67, 593-598, 1985

A:Title: Sequence analysis of a *Dictyostellium discoideum* gene coding for an active dihydro

A:Reference number: A23559; MUID:86026489

A:Accession: A23559

A:Molecule type: DNA

A:Residues: 1-369 <JAC>

A:Cross-references: GB:X02917

C:Superfamily: slime mold dihydroorotate oxidase

C:Keywords: flavoprotein; oxidoreductase; pyrimidine nucleotide biosynthesis

Query Match 40.0%; Score 42; DB 1; Length 369;

Best Local Similarity 50.0%; Pred. NO. 17;

Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 NFDFAFSIQYKKI 16

Db 19 NLENGPFIEGKKV 32

RESULT 13

S00532

gap junction protein Cx43 - rat

N:Alternate names: connexin 43; gap junction alpha-1 protein

C:Species: *Rattus norvegicus* (Norway rat)

C:Date: 30-Jun-1989 #sequence_revision 02-Jun-1994 #text_change 05-Sep-1997

C:Accession: S00532

R:Beyer, E.C.; Paul, D.L.; Goodenough, D.A.

J. Cell Biol. 105, 2621-2629, 1987

A:Title: Connexin43: a protein from rat heart homologous to a gap junction protein from

A:Reference number: S00532; MUID:88087396

A:Accession: S00532

A:Molecule type: mRNA

A:Residues: 1-382 <BEY>

A:Cross-references: EMBL:X06656; NID:g55978; PID:g55979

C:Superfamily: gap junction protein

C:Keywords: gap junction; phosphoprotein; transmembrane protein

F:1-18/Domain: intracellular #status predicted <INT1>

F:19-41/Domain: transmembrane #status predicted <INT1>

F:42-76/Domain: extracellular #status predicted <EE1>
 F:77-98/Domain: transmembrane #status predicted <TM2>
 F:99-153/Domain: intracellular #status predicted <INT2>
 F:154-184/Domain: transmembrane #status predicted <TM3>
 F:185-206/Domain: extracellular #status predicted <EE2>
 F:207-229/Domain: transmembrane #status predicted <TM4>
 F:230-382/Domain: intracellular #status predicted <INT3>
 F:297,364,365,368,369,372,373/Binding site: phosphate (Ser) (covalent) (by protein k1

Query Match 40.0%; Score 42; DB 1; Length 382;

Best Local Similarity 50.0%; Pred. NO. 18;

Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 AKNDFAPSIQYKKIA--HEL 20

Db 332 AQFFDFPDNDQNAKKAAGHEL 353

RESULT 14

D59864

hypothetical protein yktB - *Bacillus subtilis*

C:Species: *Bacillus subtilis*

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998

C:Accession: D69864

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

A:Reference number: A69580; MUID:98044033

A:Accession: D69864

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-212 <KUN>

A:Cross-references: GB:AL009126; NID:g2633699; PID:el185055; PID:g2633836

A:Experimental source: strain 168

C:Genetics:

A:Gene: yktB

Query Match 40.0%; Score 42; DB 2; Length 212;

Best Local Similarity 61.5%; Pred. NO. 9.1;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 6 FAPSIQYKKIAH 18

Db 75 FANSKRGYKKLPH 87

RESULT 15

D64433

hypothetical protein MJ1069 - *Methanococcus jannaschii*

C:Species: *Methanococcus jannaschii*

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997

C:Accession: D64433

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,

rsen, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*

A:Reference number: A64300; MUID:96337999

A:Accession: D64433
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-392 <BUL>
 A:Cross-references: GB:U67549; GB:L77117; NID:g1591709; PID:g1591721; TIGR:WJ1069; PID:g1591721
 C:Genetics:
 A:Map position: REV1010807-1009629

Query Match 40.0%; Score 42; DB 2; Length 392;
 Best Local Similarity 42.1%; Pred. No. 18;
 Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
 QY 2 KNFDFAPSIQGYKKIAHEL 20
 Db 372 ENFNWKKVKYKKIFERL 390

Search completed: November 13, 1999, 12:08:08
 Job time: 2080 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:33:44 ; Search time 51.07 Seconds
(without alignments)
11.070 Million cell updates/sec

Title: US-08-913-430-13
Perfect score: 105
Sequence: 1 AKNFDFAPSIQYKXIAHEL 20

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	46	43.8	380	1	DXAL_CHICK
2	44.5	42.4	349	1	DHFW_CAEEL
3	44	41.9	894	1	WPRA_BACSU
4	44	41.9	168	1	Y342_MYCGE
5	43.5	41.4	567	1	VE06_YACCC
6	43.5	41.4	567	1	VE06_YACCV
7	43.5	41.4	567	1	VE06_VARV
8	43	41.0	381	1	GREX_RICPR
9	43	41.0	162	1	GREX_RICPR
10	42	40.0	381	1	DXAL_RAT
11	42	40.0	369	1	DHPO_DICDI
12	41	39.0	381	1	DXAL_HUMAN
13	40	38.1	998	1	CBF_HUMAN
14	40	38.1	378	1	DXAL_XENLA
15	40	38.1	616	1	NDVA_RHIME
16	39.5	37.6	1829	1	DPOL_THEST
17	39.5	37.6	142	1	IBPB_ECOLI
18	39.5	37.6	484	1	SYE_MXCPN
19	39.5	37.6	505	1	TCMO_POPKI
20	39.5	37.6	505	1	TCMO_POPTM
21	39	37.1	1417	1	BLM_HUMAN
22	39	37.1	685	1	ERF2_YEAST
23	39	37.1	317	1	FATC_YIBAN
24	39	37.1	345	1	TMFA_TREPA
25	39	37.1	934	1	Y321_MYCGE
26	39	37.1	162	1	YHYI_ECOLI
27	38.5	36.7	314	1	DHEV_CAEEL
28	38.5	36.7	682	1	EPG_THEMA
29	38.5	36.7	664	1	YFGK_CAEEL
30	38	36.2	621	1	FOR_THELI
31	38	36.2	361	1	HIS8_HALVO
32	38	36.2	1280	1	MDR1_HUMAN
33	38	36.2	299	1	PRTM_LACLA
34	38	36.2	299	1	PRTM_LACLC
35	38	36.2	948	1	RPOP_PODAN
36	38	36.2	250	1	YK21_YEAST
37	38	36.2	213	1	YMO6_YEAST
38	37.5	35.7	896	1	CYRB_MOUSE
39	37.5	35.7	352	1	DMPF_PSESP
40	37.5	35.7	450	1	ENVZ_ECOLI
41	37.5	35.7	450	1	ENVZ_SALTI
42	37.5	35.7	450	1	ENVZ_SALTY
43	37.5	35.7	201	1	IMMU_BPSPB

44 37 35.2 796 1 ABAA_EMENI P20945 emericeella
45 37 35.2 491 1 CAFA_HAEIN P45175 haemophilus

ALIGNMENTS

```
RESULT 1
CXAL_CHICK 1
ID CXAL_CHICK STANDARD; PRT; 380 AA.
AC P14154;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GAP JUNCTION ALPHA-1 PROTEIN (CONNEXIN 43) (CX43).
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHA; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LENS;
RX MEDLINE; 90339470.
RA MUSIL L.S., BEYER E.C., GOODENOUGH D.A.;
RT "Expression of the gap junction protein connexin43 in embryonic chick
RT lens: molecular cloning, ultrastructural localization, and post-
RT translational phosphorylation.";
RL J. MEMBR. BIOL. 116:163-175(1990).
CC -|- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
CC PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXONS, THROUGH WHICH
CC MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
CC -|- SUBUNIT: A CONNEXON IS COMPOSED OF AN HEXAMER OF CONNEXINS.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -|- TISSUE SPECIFICITY: IN ALL TISSUES, BUT MOSTLY IN LENS.
CC -|- SIMILARITY: TO OTHER GAP JUNCTION PROTEINS. BELONGS TO THE ALPHA-
CC TYPE (GROUP II) SUBFAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M29003; G211643; -
CC PIR; A48171; A48171
CC PROSITE; PS00407; CONNEXINS_1; 1.
CC PFAM; PF00029; connexin; 1.
CC GAP JUNCTION; TRANSMEMBRANE.
CC INIT MET 0
CC DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 23 43 POTENTIAL.
CC DOMAIN 44 75 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 76 96 POTENTIAL.
CC DOMAIN 97 154 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 155 175 POTENTIAL.
CC DOMAIN 176 206 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 207 227 POTENTIAL.
CC DOMAIN 228 380 CYTOPLASMIC (POTENTIAL).
CC SEQUENCE 380 AA; 43041 MW; 8966C4B8 CRC32;
```

Query Match 43.8%; Score 46; DB 1; Length 380;
Best Local Similarity 54.5%; Pred. No. 3.2;
Matches 12; Conservative 2; Mismatches 6; Indels 2; Gaps 1;
QY 1 AKNFDFAPSIQYKXIA--HEL 20
|: |||| | ||: |||
Db 330 AQPFDFAEHQTKKLASGHEL 351
RESULT 2

```

DHBM_CAEEL
ID DHBM_CAEEL STANDARD; PRT; 349 AA.
AC Q17704;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PUTATIVE STEROID DEHYDROGENASE C06B3.5 (EC 1.1.1.-).
GN C06B3.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA PERCY C.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z77652; E1343767; -.
CC DR WORMPEP; C06B3.5; CE07962.
CC DR PROSITE; PS00061; ADH_SHORT; 1.
CC DR PFAM; PF00106; adh_short; 1.
CC KW HYPOTHETICAL PROTEIN; STEROID BIOSYNTHESIS; OXIDOREDUCTASE; NADP;
CC FT NP_BIND 47 76 NADP (BY SIMILARITY).
CC FT ACT_SITE 202 202 BY SIMILARITY.
CC SEQUENCE 349 AA; 39040 MW; DA240636 CRC32;
CC -----
Query Match 42.4%; Score 44.5; DB 1; Length 349;
Best Local Similarity 55.6%; Pred. No. 5.1;
Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
QY 4 FDFA-PSIOGYKKIAHEL 20
DB 105 FDETNPISIDYKRLLSQ 122
| | | | | | | | | |
RESULT 3
WPRA_BACSU STANDARD; PRT; 894 AA.
ID WPRA_BACSU
AC P54423; C06726;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CELL WALL-ASSOCIATED PROTEASE PRECURSOR (EC 3.4.21.-) [CONTAINS: CELL
DE WALL-ASSOCIATED POLYPEPTIDES CWBP23 AND CWBP52].
GN WPRA.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-54 AND 414-428.
RC STRAIN=168;
RX MARGOT P., KARAMATA D.;
RT "The wprA gene of Bacillus subtilis 168, expressed during exponential
RT growth, encodes a cell-wall-associated protease.";
RL MICROBIOLOGY 142:3437-3444(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98015415.
RA MEDINA N., VANNIER F., ROCHE B., AUTRET S., LEVINE A., SEROR S.J.;
RT "Sequencing of regions downstream of addA (98 degrees) and citG (289

```

```

degrees) in Bacillus subtilis.";
RL MICROBIOLOGY 143:3305-3308(1997).
CC -!- FUNCTION: NOT YET KNOWN; COULD BE INVOLVED IN PROTEOGLYCAN
CC DEGRADATION, BY CLEAVAGE OF ITS PEPTIDE BRIDGES.
CC -!- PTM: PROCESSED INTO CWBP23 AND CWBP52.
CC -!- SUBCELLULAR LOCATION: CELL-WALL BOUND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U58981; G1498320; -.
CC DR EMBL; Y09476; E1173511; -.
CC DR EMBL; Z99109; E1183079; -.
CC SUBTILIST; BG11846; WPRA.
CC PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC PFAM; PF00082; subtilase; 1.
CC HSP; Q99405; IMPT.
CC KW HYDROLASE; SERINE PROTEASE; CELL WALL; ZYMOCEN; SIGNAL.
CC FT SIGNAL 1 31
CC FT CHAIN 32 894 CELL WALL-ASSOCIATED PROTEASE.
CC FT CHAIN 32 ? CWBP23.
CC FT PROPEP ? 413 POTENTIAL.
CC FT CHAIN 414 894 CWBP52.
CC FT ACT_SITE 466 466 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 497 497 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 650 650 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT CONFLICT 9 9 V -> A (IN REF. 1).
CC FT CONFLICT 14 14 L -> I (IN REF. 1).
CC SEQUENCE 894 AA; 96487 MW; F46D8E72 CRC32;
CC -----
Query Match 41.9%; Score 44; DB 1; Length 894;
Best Local Similarity 38.9%; Pred. No. 17;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
QY 3 NFDFAPSIQGYKKIAHEL 20
DB 400 NYEFAPVQVEYKSLANDI 417
| | | | | | | | | |
RESULT 4
Y342_MYCGE STANDARD; PRT; 168 AA.
ID Y342_MYCGE
AC P47584; Q49340;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN MG342.
GN MG342.
OS MYCOPLASMA GENITALIUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLIICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE; 96026345.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FRITSCHMAN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITSCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL SCIENCE 270:397-403(1995).

```

```

RN  [2]
RP  SEQUENCE OF 23-128 FROM N.A.
RC  STRAIN=ATCC 33530 / G-37;
RX  MEDLINE: 94075230.
RA  PETERSON S.N., HU P.-C., BOTT K.F., HUTCHISON C.A. III:
RT  "A survey of the mycoplasma genitalium genome by using random
RT  sequencing.";
RL  J. BACTERIOL. 175:7918-7930(1993).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: U39718; G1046048;
DR  EMBL: U02231; G408972;
DR  TIGR: MG342;
KW  HYPOTHETICAL PROTEIN; ATP-BINDING.
FT  NP_BIND 119 126 ATP (POTENTIAL).
FT  CONFLICT 23 23 F -> M (IN REF. 2).
SQ  SEQUENCE 168 AA; 19128 MW; CDF9A381 CRC32;

Query Match 41.9%; Score 44; DB 1; Length 168;
Best Local Similarity 53.3%; Pred. No. 2.9;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY  2 KNFDFAPSIQYKKI 16
DB  153 ENDFKPFIEQVKKL 167

RESULT 5
VE06_VACCC STANDARD; PRT; 567 AA.
AC P21047;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE PROTEIN E6.
GN EGR.
OS VACCINIA VIRUS (STRAIN COPENHAGEN).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91021027.
RA GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PROLETTI E.;
RT "The complete DNA sequence of vaccinia virus.";
RL VIROLOGY 179:247-266(1990).
[2]
RN COMPLETE GENOME.
RP GOEBEL S.J., JOHNSON G.P., PERKUS M.E., DAVIS S.W., WINSLOW J.P.,
RA PROLETTI E.;
RL VIROLOGY 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M35027; G35392;
DR PIR: A42509; A42509.
SQ SEQUENCE 567 AA; 66669 MW; A9121B92 CRC32;

```

```

Query Match 41.4%; Score 43.5; DB 1; Length 567;
Best Local Similarity 45.0%; Pred. No. 12;
Matches 9; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY  2 KNF-DFAPSIQYKKIAHEL 20
DB  94 KNYLSFNTSIQYKTVHKL 113

RESULT 6
VE06_VACCV STANDARD; PRT; 567 AA.
AC P21607;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 01-NOV-1991 (REL. 20, LAST ANNOTATION UPDATE)
DE PROTEIN E6.
GN E6R.
OS VACCINIA VIRUS (STRAIN WR).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90377234.
RA AHN B.-Y., GERSHON P.D., JONES E.V., MOSS B.;
RT "Identification of rpo30, a vaccinia virus RNA polymerase gene with
RT structural similarity to a eucaryotic transcription elongation
RT factor.";
RL MOL. CELL. BIOL. 10:5433-5441(1990).
[2]
RN SEQUENCE FROM N.A.
RA GERSHON P.D., JONES E.V., MOSS B., AHN B.-Y.;
RL SUBMITTED (JUL-1990) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M36339; G335753;
DR PIR: E35928; E35928.
SQ SEQUENCE 567 AA; 66727 MW; EA542F73 CRC32;

Query Match 41.4%; Score 43.5; DB 1; Length 567;
Best Local Similarity 45.0%; Pred. No. 12;
Matches 9; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY  2 KNF-DFAPSIQYKKIAHEL 20
DB  94 KNYLSFNTSIQYKTVHKL 113

RESULT 7
VE06_VARV STANDARD; PRT; 567 AA.
AC P33819;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 33, LAST ANNOTATION UPDATE)
DE PROTEIN E6.
GN E6R.
OS VARIOLA VIRUS.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; POXVIRIDAE; CHORDOPOXVIRINAE;
OC ORTHOPOXVIRUS.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-INDIA-1967 / ISOLATE IND3;
RX MEDLINE: 94152154.
RA SHCHELKUNOV S.N., BLINOV V.M., RESENCHUK S.M., TOTMENIN A.V.,

```

RA SANDAKHCHIEV L.S.;
 RT "Analysis of the nucleotide sequence of a 43 kbp segment of the
 RL genome of variola virus India-1967 strain.";
 RN VIRUS RES. 30:239-258(1993).
 RP COMPLETE GENOME.
 RC STRAIN=INDIA-1967 / ISOLATE IND3;
 RX MEDLINE; 93202281.
 RA SCHELKUNOV S.N., BLINOV V.M., SANDAKHCHIEV L.S.;
 RT "Genes of variola and vaccinia viruses necessary to overcome the host
 RL protective mechanisms.";
 RN FEBS LETT. 319:80-83(1993).
 RP SEQUENCE FROM N.A.
 RC STRAIN=BANGLADESH-1975;
 RX MEDLINE; 94088747.
 RA MASSUNG R.F., ESPOSITO J.J., LIU L., QI J., UTERBACK T.R.,
 RA KNIGHT J.C., AUBIN L., YURAN T.E., PARSONS J.M., LOPAREV V.N.,
 RA SELIVANOV N.A., CAVALLARO K.F., KERLAVAGE A.R., MAHY B.W.J.,
 RA VENTER C.J.;
 RT "Potential virulence determinants in terminal regions of variola
 RL smallpox virus genome";
 RN NATURE 366:748-751(1993).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X69198; G297228; -
 DR EMBL; L22579; G438965; -
 DR PIR; I36841; I36841.
 SQ SEQUENCE 567 AA; 66717 MW; F4D051C8 CRC32;
 Query Match 41.4%; Score 43.5; DB 1; Length 567;
 Best Local Similarity 45.0%; Pred. No. 12;
 Matches 9; Conservative 5; Mismatches 5; Indels 1; Gaps 1;
 QY 2 KNF-DFAPSIQGYKKIAHEL 20
 ||: ||: ||: ||:
 Db 94 KNYLSFSNAIQSYKETHKL 113
 RESULT 8
 ID CXAL_MOUSE STANDARD; PRT; 381 AA.
 AC P23242;
 DT 01-NOV-1991 (REL. 20, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE GAP JUNCTION ALPHA-1 PROTEIN (CONNEXIN 43) (CX43) (GAP JUNCTION 43 KD
 DE HEART PROTEIN).
 GN GJAL OR CXN-43.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 91217014.
 RA BEYER E.C., STEINBERG T.H.;
 RT "Evidence that the gap junction protein connexin-43 is the
 RL ATP-induced pore of mouse macrophages.";
 RN J. BIOL. CHEM. 266:7971-7974(1991).
 RP SEQUENCE FROM N.A.
 RA RUANGVORAVAT C.P., MORGAN J.L., LO C.W.;
 RL SUBMITTED (SEP-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.

RC TISSUE=OVARY;
 RX MEDLINE; 91285228.
 RA NISHI M., KUMAR N.M., GILULA N.B.;
 RT "Developmental regulation of gap junction gene expression during
 RL mouse embryonic development.";
 RN DEV. BIOL. 146:117-130(1991).
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE; 92299686.
 RA HENNEMANN J., SUCHYNA T., LICHTENBERG-FRATHE H., JUNGBLUTH S.,
 RA DAHL E., SCHWARZ J., NICHOLSON B.J., WILLECKE K.;
 RT "Molecular cloning and functional expression of mouse connexin40, a
 RL second gap junction gene preferentially expressed in lung.";
 RN J. CELL BIOL. 117:1299-1310(1992).
 CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
 CC PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXONS, THROUGH WHICH
 CC MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
 CC -1- FUNCTION: CONNEXIN 43 IS POSSIBLY THE ATP-INDUCED PORE OF
 CC MOUSE MACROPHAGES.
 CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF AN HEXAMER OF CONNEXINS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: TO OTHER GAP JUNCTION PROTEINS. BELONGS TO THE ALPHA-
 CC TYPE (GROUP II) SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X61896; G192679; -
 DR EMBL; X61576; G50510; -
 DR EMBL; X63801; G567197; -
 DR EMBL; X62836; G50523; -
 DR PIR; A39802; A39802.
 DR PIR; S18110; S18110.
 DR PIR; S23110; S23110.
 DR MGD; MGI:95713; GJAL.
 DR PROSITE; PS00407; CONNEXINS_1; 1.
 DR PROSITE; PS00408; CONNEXINS_2; 1.
 DR PFAM; PF00029; connexin; 1.
 KW GAP JUNCTION; TRANSMEMBRANE.
 FT INIT_MET 0 0
 FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 23 43 POTENTIAL.
 FT DOMAIN 44 75 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 76 96 POTENTIAL.
 FT DOMAIN 97 154 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 155 175 POTENTIAL.
 FT DOMAIN 176 206 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 207 227 POTENTIAL.
 FT DOMAIN 228 381 CYTOPLASMIC (POTENTIAL).
 FT CONFLICT 319 319 M -> T (IN REF. 3).
 SQ SEQUENCE 381 AA; 42873 MW; 3B503DEC CRC32;
 Query Match 41.0%; Score 43; DB 1; Length 381;
 Best Local Similarity 50.0%; Pred. No. 9;
 Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;
 QY 1 AKNFDAPSIQGYKKIA--HEL 20
 |: ||: | ||: ||:
 Db 331 AQPFDPPDSQNAKVAAGHEL 352
 RESULT 9
 ID GREAR_RICPR STANDARD; PRT; 162 AA.
 AC P27640;
 DT 01-AUG-1992 (REL. 23, CREATED)

DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE TRANSCRIPTION ELONGATION FACTOR GRETA (TRANSCRIPT CLEAVAGE FACTOR
 DE GRETA).
 GN RP861 OR GRETA.
 OS RICKETTSIA PROWAZEKII.
 OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RICKETTSIALES;
 OC RICKETTSIAEAE; RICKETTSIAE; RICKETTSIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MADRID E.
 RX MEDLINE: 92350688.
 RA MARKS G.L., WOOD D.O.;
 RL "Nucleotide sequence of the Rickettsia prowazekii grea homolog.";
 RL NUCLEIC ACIDS RES. 20:3785-3785(1992).
 CC -1- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION
 CC ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING
 CC SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION
 CC OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN
 CC LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY
 CC CLEAVAGE FACTORS SUCH AS GRETA OR GREB ALLOWS THE RESUMPTION OF
 CC ELONGATION FROM THE NEW 3'TERMINUS. GRETA RELEASES SEQUENCES OF
 CC 2 TO 3 NUCLEOTIDES.
 CC -1- SIMILARITY: BELONGS TO THE GRE/GREB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z12122; G46346; -
 CC EMBL: U02878; G434678; -
 CC PIR: S21472; S21472.
 CC PIR: S26176; S26176.
 CC PROSITE: PS00829; GREAB.1; 1.
 CC PROSITE: PS00830; GREAB.2; 1.
 CC PIR: PF01372; GREa_Greb; 1.
 CC HSP: P21346; IGRJ.
 KW TRANSCRIPTION REGULATION; DNA-BINDING.
 SQ SEQUENCE 162 AA; 18230 MW; 77C6BF27 CRC32;

Query Match 41.0%; Score 43; DB 1; Length 162;
 Best Local Similarity 44.4%; Pred. No. 4;
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 3 NFDPAFSGYKKIAHEL 20
 I : : : : :
 Db 2 NTRFPITANGFKKLEHEL 19

RESULT 10
 CXAL_RAT
 ID CXAL_RAT STANDARD; PRT; 381 AA.
 AC P08050;
 DT 01-AUG-1988 (REL. 08, CREATED)
 DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE GAP JUNCTION ALPHA-1 PROTEIN (CONNEXIN 43) (CX43) (GAP JUNCTION 43 KD
 DE HEART PROTEIN).
 GN GJAL OR CXN-43.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 88087396.
 RA BEYER E.C., PAUL D.L., GOODENOUGH D.A.;
 RT "Connexin43: a protein from rat heart homologous to a gap junction
 RT protein from liver.";

RL J. CELL BIOL. 105:2621-2629(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=UTERUS;
 RX MEDLINE: 91241281.
 RA LANG L.M., BEYER E.C., SCHWARTZ A.L., GITLIN J.D.;
 RT "Molecular cloning of a rat uterine gap junction protein and analysis
 RT of gene expression during gestation."
 RL AM. J. PHYSIOL. 260:E787-E793(1991).
 RN [3]
 RP SEQUENCE OF 1-32.
 RC TISSUE=HEART;
 RX MEDLINE: 95207650.
 RA NICHOLSON B.J., GROS D.B., KENT S.B.H., HOOD L.E., REVEL J.-P.;
 RT "The Mr 28,000 gap junction proteins from rat heart and liver are
 RT different but related."
 RL J. BIOL. CHEM. 260:6514-6517(1985).
 RN [4]
 RP PARTIAL SEQUENCE OF 1-15.
 RC TISSUE=BRAIN;
 RX MEDLINE: 91348048.
 RA DUPONT E., EL AOUARI A., FROMAGET C., BRIAND J.-C., GROS D.;
 RT "Affinity purification of a rat-brain junctional protein, connexin
 RT 43."
 RL EUR. J. BIOCHEM. 200:263-270(1991).
 RN [5]
 RP DISULFIDE BONDS.
 RX MEDLINE: 91337077.
 RA JOHN S.A., REVEL J.-P.;
 RT "Connexon integrity is maintained by non-covalent bonds:
 RT intramolecular disulfide bonds link the extracellular domains in rat
 RT connexin-43."
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 178:1312-1318(1991).
 RN [6]
 RP TOPOLOGY.
 RX MEDLINE: 92167270.
 RA YEAGER M., GIULIA N.B.;
 RT "Membrane topology and quaternary structure of cardiac gap junction
 RT ion channels."
 RL J. MOL. BIOL. 223:929-948(1992).
 CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
 CC PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXONS, THROUGH WHICH
 CC MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
 CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF AN HEXAMER OF CONNEXINS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- PTM: THERE IS AT LEAST ONE INTRAMOLECULAR DISULFIDE BOND.
 CC -1- SIMILARITY: TO OTHER GAP JUNCTION PROTEINS. BELONGS TO THE ALPHA-
 CC TYPE (GROUP II) SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M19317; G203507; -
 CC EMBL: X06856; G55979; -
 CC PIR: S00532; S00532.
 CC PIR: A24047; A24047.
 CC PROSITE: PS00407; CONNEXINS_1; 1.
 CC PROSITE: PS00408; CONNEXINS_2; 1.
 CC PFAM: PF00029; connexin; 1.
 KW GAP JUNCTION; TRANSMEMBRANE.
 FT INIT_MET 0
 FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 23 43 POTENTIAL.
 FT DOMAIN 44 75 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 76 96 POTENTIAL.
 FT DOMAIN 97 154 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 155 175 POTENTIAL.
 FT DOMAIN 176 206 EXTRACELLULAR (POTENTIAL).


```
FT TRANSNM 207 227 POTENTIAL.
FT DOMAIN 228 381 CYTOPLASMIC (POTENTIAL).
FT CONFLICT 1 1 G -> A (IN REF. 3).
FT CONFLICT 15 15 A -> T (IN REF. 2).
FT CONFLICT 27 27 V -> I (IN REF. 3).
SQ SEQUENCE 381 AA; 42900 MW; 69B1DF0 CRC32;

Query Match 40.0%; Score 42; DB 1; Length 381;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 AKNDFAPSIQYKKIA--HEL 20
   : ||| | ||| |||
Db 331 AQPDPDDNQNAKKAAGHEL 352

RESULT 11
DHDO_DICDI
ID DHDO_DICDI STANDARD; PRT; 369 AA.
AC P07670;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE DIHYDROOROTATE DEHYDROGENASE (EC 1.3.99.11).
GN PYR4.
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86026489.
RA JACQUET M., KALEKINE M., BOY-MARCOTTE E.;
RT "Sequence analysis of a Dictyostelium discoideum gene coding for an
RT active dihydroorotate dehydrogenase in yeast.";
RL BIOCHIMIE 67:583-588(1985).
CC -1- CATALYTIC ACTIVITY: L-DIHYDROOROTATE + ACCEPTOR -> OROTATE +
CC REDUCED ACCEPTOR.
CC -1- PATHWAY: FOURTH STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -1- SIMILARITY: NO DIRECT SEQUENCE HOMOLOGY CAN BE DETECTED AMONG THE
CC DIHYDROOROTATE OXIDASE (EC 1.3.3.1) (PYRD/URAL GENE) AND THIS
CC ENZYME.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02917; G7253; ALT_SEQ.
DR PIR; A23559; A23559.
DR DICTYDB; DD05000; PYR4.
DR PFAM; PF01180; Dhodehase; 1.
KW PYRIMIDINE BIOSYNTHESIS; OXIDOREDUCTASE.
SQ SEQUENCE 369 AA; 40370 MW; 102BA697 CRC32;

Query Match 40.0%; Score 42; DB 1; Length 369;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 NFDAPSIQYKKI 16
   : | | | | |
Db 19 NLENGPIEGYKKV 32

RESULT 12
CXAL_HUMAN
ID CXAL_HUMAN STANDARD; PRT; 381 AA.
AC P17302;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)

01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
GAP JUNCTION ALPHA-1 PROTEIN (CONNEXIN 43) (CX43) (GAP JUNCTION 43 KD
HEART PROTEIN).
GN GJAL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91257836.
RA FISHMAN G.I., EDDY R.L., SHOWS T.B., ROSENTHAL L., LEINWAND L.A.;
RT "The human connexin gene family of gap junction proteins: distinct
RT chromosomal locations but similar structures.";
RL GENOMICS 10:250-256(1991).
CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
CC PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXONS, THROUGH WHICH
CC MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF AN HEXAMER OF CONNEXINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: TO OTHER GAP JUNCTION PROTEINS. BELONGS TO THE ALPHA-
CC TYPE (GROUP II) SUBFAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52947; G29917;
DR EMBL; M65188; G181209;
DR PIR; S10470; S10470.
DR PIR; A35853; A35853.
DR MIM; 121014;
DR PROSITE; PS00407; CONNEXINS_1; 1.
DR PROSITE; PS00408; CONNEXINS_2; 1.
DR PFAM; PF00029; connexin; 1.
KW GAP JUNCTION; TRANSMEMBRANE.
FT INIT_MET 0 0
FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
FT TRANSNM 23 43 POTENTIAL.
FT DOMAIN 44 75 EXTRACELLULAR (POTENTIAL).
FT TRANSNM 76 96 POTENTIAL.
FT DOMAIN 97 154 CYTOPLASMIC (POTENTIAL).
FT TRANSNM 155 175 POTENTIAL.
FT DOMAIN 176 206 EXTRACELLULAR (POTENTIAL).
FT TRANSNM 207 227 POTENTIAL.
FT DOMAIN 228 381 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 381 AA; 42877 MW; 039AB1A5 CRC32;

Query Match 39.0%; Score 41; DB 1; Length 381;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 AKNDFAPSIQYKKIA--HEL 20
   : ||| | ||| |||
Db 331 AQPDPDDNQNAKKAAGHEL 352

RESULT 13
CBF_HUMAN
ID CBF_HUMAN STANDARD; PRT; 998 AA.
AC Q03701;
```

modified and this statement is not removed. Usage by and for commercial/CC
entitles requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).-----CC

EMBL; X17243; G64499; -
PIR; A34575; A34575.
PROSITE; PS00407; CONNEXINS_1; 1.
DR DR PROSITE; PS00408; CONNEXINS_2; 1.
DR DR PFAM; PF00029; connexin; 1.
DR DR GAP JUNCTION; TRANSMEMBRANE.
FF FT INIT MET 0 0
FF FT DOMAIN 1 22 BY SIMILARITY
FF FT TRANSMEM 23 43 CYTOPLASMIC (POTENTIAL).
FF FT DOMAIN 44 75 EXTRACELLULAR (POTENTIAL).
FF FT TRANSMEM 76 96 POTENTIAL.
FF FT DOMAIN 97 154 CYTOPLASMIC (POTENTIAL).
FF FT TRANSMEM 155 175 POTENTIAL.
FF FT DOMAIN 176 206 EXTRACELLULAR (POTENTIAL).
FF FT TRANSMEM 207 227 POTENTIAL.
FF FT DOMAIN 228 378 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 378 AA; 42830 MW; 18E9A5AB CRC32;

Query Match 38.1%; Score 40; DB 1; Length 378;
Best Local Similarity 45.5%; Pred. No. 31;
Matches 10; Conservative 4; Mismatches 6; Indels 2; Gaps 1;

QY 1 AKNFDFAPSIQGYKKIA--HEL 20
| : ||| : | || : || :
DB 328 AQPFDSDEHQNTKMPGHEM 349

RESULT 15
NDVA_RHIME
ID NDVA_RHIME STANDARD; PRT; 516 AA.
PI8767;
AC 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DT BETA-(1->2)GLUCAN EXPORT ATP-BINDING PROTEIN NDVA.
GN NDVA.
OS RHIZOBIMUM MELIOTTI.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
RHIZOBIACEAE; SINORHIZOBIUM.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 88298659.
RX STANFIELD S.W., IELPI L., O'BROCHTA D., HELINSKI D.R., DITTA G.S.;
RA "The ndva gene product of Rhizobium meliloti is required for
RT beta-(1->2)glucan production and has homology to the ATP-binding
RT export protein HlyB";
RL J. BACTERIOL. 170:3523-3530(1988).
CC -!- FUNCTION: INVOLVED IN THE EXPORT OF BETA-(1,2)-GLUCAN WHICH ARE
CC REQUIRED FOR NODULATION OF LEGUME ROOTS.
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS).
-----CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial/
entitles requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).-----CC

EMBL; M20726; G152269; -
PIR; A31094; VKZRNA.
DR DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR DR PFAM; PF00005; ABC_tran; 1.
DR DR PFAM; PF00664; ABC_membrane; 1.
DR DR HSP; PI3569; INBD.
FT SUGAR TRANSPORT; TRANSPORT; ATP-BINDING; MODULATION.
FT NP_BIND 399 406 ATP (BY SIMILARITY).

SQ SEQUENCE 616 AA: 67238 MW: 6E284B5B CRC32:

Query Match 38.1%; Score 40; DB 1; Length 616;
Best Local Similarity 46.7%; Pred. No. 51;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 NFDFAPSIQGYKIA 17
:|||||: : :
Db 372 SFDFANSAQGVNRVS 386

Search completed: November 13, 1999, 10:33:45
Job time: 5195 sec

GenCore version 4.5

Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:39 ; Search time 139.86 Seconds
(without alignments)
8.801 Million cell updates/sec

Title: US-08-913-430-13

Perfect score: 105

Sequence: 1 AKNDFAFSIQGYKKIAHEL 20

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SP TREMBL_10.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	91.4	627	2	O30643
2	96	91.4	625	2	O30704 mycoplasma
3	45	42.9	682	2	O51192 borrelia bu
4	45	42.9	370	3	O94026 candida alb
5	45	42.9	88	10	O81379
6	44	41.9	211	2	O67841 aquifex aeo
7	44	41.9	567	5	O77349 plasmidium
8	43.5	41.4	567	12	O57189 vaccinia v1
9	42	40.0	392	1	O58469 methanococ
10	42	40.0	201	1	O52460
11	42	40.0	1358	2	P96967
12	42	40.0	224	2	O06076 mycobacteri
13	42	40.0	212	2	O45498 bacillus su
14	42	40.0	1723	2	P72194 porphyromon
15	42	40.0	436	3	O07794 saccharomyc
16	41	39.0	575	2	O51775 pseudomonas
17	41	39.0	497	2	P74864 salmonella
18	41	39.0	240	3	O08954
19	41	39.0	405	3	O59702 schizosacch
20	40.5	38.6	337	2	O34232 vibrio chol
21	40.5	38.6	337	2	O87136 vibrio chol
22	40	38.1	865	1	O27595 methanobact
23	40	38.1	358	1	O58692 pyrococcus
24	40	38.1	613	2	O45882 clostridium
25	40	38.1	1767	5	O19346 caenorhabdi
26	39.5	37.6	396	10	O40907 populus kit
27	39	37.1	308	2	O25411 helicobacte
28	39	37.1	687	2	O83373 proteus mir
29	39	37.1	308	2	O921d4 helicobacte

30	39	37.1	224	3	O07089
31	39	37.1	162	3	O14190
32	39	37.1	2609	3	O74390
33	39	37.1	1004	4	P78364
34	39	37.1	948	4	O15290
35	39	37.1	1078	4	O95486
36	39	37.1	312	5	O20029
37	39	37.1	2247	5	O21480
38	39	37.1	4767	5	O17301
39	39	37.1	2391	5	O27732
40	39	37.1	362	5	O62142
41	39	37.1	197	9	O38147
42	39	37.1	507	10	O81492
43	39	37.1	470	10	O92QP4
44	39	37.1	1012	11	O64028
45	39	37.1	1587	11	P70287

ALIGNMENTS

RESULT 1
O30643
ID O30643 PRELIMINARY; PRT: 627 AA.
AC O30643;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE PROLIPOPROTEIN P65 PRECURSOR.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J, ATCC25934;
RA HEIDARI M.B., KIM M.F., MCINTOSH M.A., WISE K.W.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF013714; AAB67173.1; -
DR PFAM; PF00657; Lipase_GDSL; 1.
KW Signal; Lipoprotein.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 627 LIPOPROTEIN P65.
SQ SEQUENCE 627 AA; 71016 MW; 10C5A2A5 CRC32;

Query Match 91.4%; Score 96; DB 2; Length 627;
Best Local Similarity 90.0%; Pred. No. 9e-08;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKNDFAFSIQGYKKIAHEL 20
Db 318 AKNDFAFSIQGYKKIAHQL 337
||||| |||||||||

RESULT 2
O30704
ID O30704 PRELIMINARY; PRT: 625 AA.
AC O30704;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE SURFACE LIPOPROTEIN PRECURSOR.
OS Mycoplasma hyopneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-232A;
RA HSU T., MINION F.C.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF015665; AAB70214.1; -
DR PFAM; PF00657; Lipase_GDSL; 1.
KW Signal; Lipoprotein.

FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 625 MATURE SURFACE LIPOPROTEIN.
SQ SEQUENCE 625 AA; 70914 MW; C109763D CRC32;

Query Match 91.4%; Score 96; DB 2; Length 625;
Best Local Similarity 90.0%; Pred. No. 9e-08;
Matches 18; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKNDFAPSIQGYKKIAHEL 20

||||| ||||| ||||| |||||

Db 318 AKNDFHPSIQGYKKIAHQL 337

RESULT 3

OS1192 ID OS1192 PRELIMINARY; PRT; 682 AA.

AC OS1192; 1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE HYPOTHETICAL 78.5 KD PROTEIN.

GN BB0170.

OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 35210 / B31;

RX MEDLINE; 98065943.

RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,

RA LATHGIRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,

RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,

RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,

RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,

RA UTTERBACK T., WATHEY L., McDONALD L., ARTIACH P., BOWMAN C.,

RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,

RA SMITH H.O., VENTER J.C.;

RT "Genomic sequence of a Lyme disease spirochaete, Borrelia

RT burgdorferi."

RL Nature 390:580-586(1997).

RL EMBL; AE001128; AAC66568.1;

DR TIGR; BB0170;

KW Hypothetical protein.

SQ SEQUENCE 682 AA; 78474 MW; 4233033A CRC32;

Query Match 42.9%; Score 45; DB 2; Length 682;

Best Local Similarity 61.3%; Pred. No. 24;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 KNDFAPSIQGYK 14

||| ||| |||

Db 317 KNYSFKSKNGYK 329

RESULT 4

ID OS4026 PRELIMINARY; PRT; 370 AA.

AC OS4026; 1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)

DE HYPOTHETICAL 42.5 KD PROTEIN.

GN CA49C10.12C.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;

OC Candidaceae; Candida.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-1161;

RA TAYLOR K., HARRIS D.;

RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-1161;
RA BARRELL B.G., RAJANDREAM M.A.;

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-1161;

RA TAIT E., SIMON M.C., KING S., BROWN A.J., GOW N.A.R., SHAW D.J.;

RT "A Candida albicans Genome Project: Cosmid Contigs, Physical Mapping,

RT and Gene Isolation."

RL Fungal Genet. Biol. 21:308-314(1997).

DR EMBL; AL033497; CAA21976.1;

KW Hypothetical protein.

SQ SEQUENCE 370 AA; 42505 MW; B47615FC CRC32;

Query Match 42.9%; Score 45; DB 3; Length 370;

Best Local Similarity 45.0%; Pred. No. 12;

Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 AKNDFAPSIQGYKKIAHEL 20

||||| ||| |||

Db 171 SKNIDFGYLVDAKKFMHEL 190

RESULT 5

OS1379 ID OS1379 PRELIMINARY; PRT; 88 AA.

AC OS1379; 1998 (TREMBlrel. 08, Created)

DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE REVERSE TRANSCRIPTASE (FRAGMENT).

GN RT.

OS Lycopersicon esculentum (Tomato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;

OC Solanum.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. H722; TISSUE-LEAF;

RA ROGERS S.A., PAULS K.P.;

RT "Tyl-copia class retrotransposons of tomato."

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF072646; AAC34608.1;

KW RNA-directed DNA polymerase.

FT NON_TER 1

FT NON_TER 88

SQ SEQUENCE 88 AA; 10440 MW; 42C08971 CRC32;

Query Match 42.9%; Score 45; DB 10; Length 88;

Best Local Similarity 44.4%; Pred. No. 2.6;

Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 KNDFAPSIQGYKKIAHE 19

||||| ||||| |||||

Db 48 KRFEYVMGEQGYKKISSD 65

RESULT 6

OS67841 ID OS67841 PRELIMINARY; PRT; 211 AA.

AC OS67841; 1998 (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-AUG-1998 (TREMBlrel. 08, Last annotation update)

DE 2-ACYLGLYCEROPHOSPHOETHANOLAMINE ACYLTRANSFERASE.

GN AAS.

OS Aquifex aeolicus.

OC Bacteria; Aquificales; Aquificaceae; Aquifex.

RN [1]

RESULT	8
Q57189	
ID	C57189 PRELIMINARY; PRT; 567 AA.
AC	C57189;
DT	01-JUN-1998 (TREMBLrel. 06, Created)
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT	01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE	PURATIVE 66.7K PROTEIN.
MVA053R.	GN MVA053R.
OS	Vaccinia virus.
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC	Orthopoxvirus.
[1]	
RN	SEQUENCE FROM N.A.
RP	

RESULT	10	
OS9460		
ID	059460	PRELIMINARY; PRT; 201 AA.
AC	059460;	
DT	01-AUG-1998 (TReMBLrel. 07, Created)	
DT	01-AUG-1998 (TReMBLrel. 07, last sequence update)	
DT	01-JAN-1999 (TReMBLrel. 09, last annotation update)	
DE	201AA LONG HYPOTHETICAL PROTEIN.	
PI	1796.	
OS	Pyrococcus horikoshii.	
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-OT3;	
RX	MEDLINE: 98344137.	

RP SEQUENCE FROM N.A.
RX MEDLINE; 93188700.
RA EIGLMETTER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
RT of *Mycobacterium leprae*.";
RL MBL. Microbiol. 7:197-206(1993).
DR EMBL. 295398; CAB08810.1; -.
LR Hypothetical protein.
SQ SEQUENCE 224 AA; 24176 MW; 9ADC21AC CRC32;

Query Match 40.0%; Score 42; DB 2; Length 224;
Best Local Similarity 38.9%; Pred. No. 22;
Matches 7; Conservative 5; Mismatches 6; Indels

```
QY      3 NFDFAPSIQGYKKIAHEL 20
      | | | : | | | : |
Db     120 NTDEDAAVQPYRKVVEKL 137
```

RESULT	13
Q45498	
ID	PRELIMINARY; PRT: 212 AA.
AC	Q45498;
DT	01-NOV-1996 (TREMBlrel. 01, Created)
DT	01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT	01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE	HYPOTHETICAL 24.6 KD PROTEIN.
DE	YKTB.
GN	Bacillus subtilis.
OS	Bacillus subtilis.
OC	Bacteria: Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
RN	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=I168;
RA	MEDLINE; 90368558.
RX	HEMILA H., PALVA A., PAULIN L., ARVIDSON S., PALVA I.;
RT	"Secretory S complex of Bacillus subtilis: sequence analysis and
RL	identity to pyruvate dehydrogenase.";
RL	J. Bacteriol. 172:5052-5063(1990).
RT	[2]
RN	SEQUENCE FROM N.A.
RP	RP

KN	SEQUENCE FROM N.A.
RP	
RC	STRAIN-1168;
RX	MEDLINE; 97124187.
RA	WINTERS P., CALDWELL R., ENFIELD L., FERRARI E.;
RT	"The ampS-prpE (124 degrees-127 degrees) region of the Bacillus
RT	subtilis 168 chromosome: sequencing of a 27 kb segment and
RT	identification of several genes in the area.";
RL	Microbiology 142:0-0(0).

"TRAKS-2"; CARROLL, C., EARLE, D., FARRAR, G.,
RT
The amp-nprE (124 degrees-17 degrees) region of the Bacillus
RT
subtilis i68 chromosome: sequencing of a 27 kb segment and
RT
RT identification of several genes in the area."
RL Microbiology 142:0-0(0).
[3]
RN

[3]
KN SEQUENCE FROM N.A.
RP STRAIN-1168;
RX MEDLINE; 97144523.
RA HENRIQUES A.O., BRYAN E.M., BEALL B.W., MORAN C.P. JR.;
RT "cse15, cse60, and csr22 are new members of mother-cell-specific
RL sporulation regulons in *Bacillus subtilis*.";
RT J. Bacteriol. 179:389-398(1997).

SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE; 98044033.
RX KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHIERT S.,
RA BORRILLER S., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA DENTZOT F., DEVINE K.M., DUSTERHOFT A., EHRICH S.D., EMERSON P.T.,
RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA GHIM S.Y., GLASER P., GORFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA GUSEPPI G., GUY B.J., HAGA K., HATECH J., HARWOOD C.R., HENNAUT A.,
RA HILBERT H., HOLSAPPEL S., HOSONO S., HULIO M.F., ITAYA M., JONES L.,
RA JORTIS B., KARAMATA D., KASAHARA Y., KLAUSER-BLANCHARD M., KLEIN C.,

RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
 RA KURIITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 RA NOONE D., O'REILLY M., OGAWA K., OGAWA K., OUDEGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTELELLA D., PORWOLLIK S., PRESCOTT A.M.,
 RA PRESKAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
 RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDO B.,
 RA SOROKIN A., TACCINI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
 RA VIARI A., WAMBUUT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*
 RT *subtilis*,"
 RL Nature 390:249-256(1997).
 DR EMBL; AF012285; AAC24939.1; -.
 DR EMBL; Z99111; CAB13338.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 212 AA; 24616 MW; 824502E2 CRC32;

Query Match 40.0%; Score 42; DB 2; Length 212;
 Best Local Similarity 61.5%; Pred. No. 21;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 6 FAPSIQYKKIAH 18

| | | : ||||: |

Db 75 FANSKRGYKKLPH 87

RESULT 14

P72194 PRELIMINARY; PRT; 1723 AA.

AC P72194; PRT; 1723 AA.

DT 01-FEB-1997 (Tremblrel. 02, Created)

DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)

DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)

DE LYS-GINGIPAIN.

GN KGP.

OS Porphyromonas gingivalis (Bacteroides gingivalis).

OC Bacteria; Cytophagales; Bacteroidaceae; Porphyromonas.

RN [1]

RC SEQUENCE FROM N.A.

RX STRAIN-381;

RX MEDLINE; 97044756.

RA OKAMOTO K., KADOWAKI T., NAKAYAMA K., YAMAMOTO K.,

RT "Cloning and sequencing of the gene encoding a novel lysine-specific

RT cysteine proteinase (Lys-gingipain) in *Porphyromonas gingivalis*:

RT structural relationship with the arginine-specific cysteine

RT proteinase (Arg-gingipain).";

RL J. Biochem. 120:398-406(1996).

DR EMBL; D83258; BAA11870.1; -.

DR PFAM; PF01364; Peptidase_C25; 2.

SQ SEQUENCE 1723 AA; 187361 MW; B589E9D6 CRC32;

Query Match 40.0%; Score 42; DB 2; Length 1723;
 Best Local Similarity 38.9%; Pred. No. 2e+02;
 Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFDFAPSIQYKKIAHEL 20

| | | : ||||: ||

Db 1495 NFLITPKVGAKKITYKV 1512

RESULT 15

Q07794

ID Q07794

AC Q07794;

PRELIMINARY; PRT; 436 AA.

PRT; 436 AA.

PRT; 436 AA.

PRT; 436 AA.

PRT; 436 AA.

PRT; 436 AA.

PRT; 436 AA.

PRT; 436 AA.

DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE CHROMOSOME XII READING FRAME ORF YLL002W.
 GN L1377.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MIOGA T., ZIMMERMANN F.K.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MIPS;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FY23 /RD005 (DERIVED FROM S288C);
 RX MEDLINE; 96405918.
 RA MIOGA T., ZIMMERMANN F.K.;
 RT "Sequence analysis of the CEN12 region of *Saccharomyces cerevisiae* on
 RT a 43.7 kb fragment of chromosome XII including an open reading frame
 RT homologous to the human cystic fibrosis transmembrane conductance
 RT regulator protein CFTR.";
 RL Yeast 12:693-708(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FY23 /RD005 (DERIVED FROM S288C);
 RX MEDLINE; 91094833.
 RA DAMAK F., BOY-MARCOTTE E., LE-ROSCOUET D., GUILBAUD R., JACQUET M.;
 RT "SDC25, a CDC25-like gene which contains a RAS-activating domain and
 RT is a dispensable gene of *Saccharomyces cerevisiae*.";
 RL Mol. Cell. Biol. 11:202-212(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FY23 /RD005 (DERIVED FROM S288C);
 RX MEDLINE; 93087480.
 RA RIPAUMASTER T.L., VAUGHN G.P., WOOLFORD J.L. JR.;
 RT "A putative ATP-dependent RNA helicase involved in *Saccharomyces*
 RT *cerevisiae* ribosome assembly.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135(1992).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FY23 /RD005 (DERIVED FROM S288C);
 RX MEDLINE; 94375516.
 RA BURGESS S.W., DELANNOY M., JENSEN R.E.;
 RT "MMW1 encodes a mitochondrial outer membrane protein essential for
 RT establishing and maintaining the structure of yeast mitochondria.";
 RL J. Cell Biol. 126:1375-1391(1994).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FY23 /RD005 (DERIVED FROM S288C);
 RX MEDLINE; 95348179.
 RA GAMMIE A.E., KURIHARA L.J., VALLEE R.B., ROSE M.D.;
 RT "DNM1, a dynamin-related gene, participates in endosomal trafficking
 RT in yeast.";
 RL J. Cell Biol. 130:553-566(1995).
 DR EMBL; Z73107; CAA97445.1; -.
 DR EMBL; X91488; CAA62768.1; -.
 SQ SEQUENCE 436 AA; 50095 MW; 3249F9CD CRC32;

Query Match 40.0%; Score 42; DB 3; Length 436;
 Best Local Similarity 61.5%; Pred. No. 45;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 PSIOGYKKIAHEL 20

| | | : ||||: ||

Db 125 PAIRSYKKISPTEL 137

Search completed: November 13, 1999, 12:55:40
Job time: 3039 sec

Result No.	Score	Query			DB	ID	Description
		Match	Length	%			
1	59	100.0	12	1	W01026	Mycoplasma 72-75 k	
2	53	89.8	627	1	W62451	Mycoplasma hyopneumoniae	
3	36	61.0	3135	1	RS7474	P. faicai parum transmembrane protein	
4	36	61.0	579	1	W31365	C16N for promoting cell growth	
5	36	61.0	579	1	W31366	C16N for promoting cell growth	
6	35	59.3	567	1	R36601	TGF-beta receptor	
7	35	59.3	1068	1	R43342	Human p110 Receptor	
8	35	59.3	1068	1	R43341	p110. Recombinant	
9	35	59.3	567	1	R88215	Human TGF-beta type II receptor	
10	35	59.3	432	1	W20733	H. pylori cell envelope	
11	35	59.3	100	1	W20358	H. pylori cell envelope	
12	34	57.6	304	1	P94623	Porin of clone P1	
13	34	57.6	492	1	P94263	Bovine adrenal gland	
14	34	57.6	304	1	P70637	Sequence encoded by human cadherin-12	
15	34	57.6	794	1	W13135	Putative human cadherin-12	
16	34	57.6	794	1	W25637	Human cadherin-12	
17	33	55.9	877	1	R28348	Bacillus caldota flagellin, New flagellin	
18	33	55.9	272	1	R38791	B. stearothermophil flagellin, New flagellin	
19	33	55.9	876	1	R45155	B. stearothermophil flagellin, New flagellin	
20	33	55.9	693	1	R83122	Human guanosine 5' triphosphate	
21	33	55.9	693	1	R83123	Human guanosine 5' triphosphate	
22	33	55.9	876	1	R80137	B. stearothermophil flagellin, New flagellin	
23	33	55.9	876	1	R80142	B. stearothermophil flagellin, New flagellin	
24	33	55.9	876	1	R80143	B. stearothermophil flagellin, New flagellin	
25	33	55.9	433	1	W00365	Human cyclin B1, D	
26	33	55.9	877	1	W22845	Bacillus caldota flagellin, New flagellin	
27	33	55.9	877	1	W22846	Bacillus caldota flagellin, New flagellin	
28	33	55.9	877	1	W22847	Bacillus caldota flagellin, New flagellin	
29	33	55.9	876	1	W35905	Bacillus caldota flagellin, New flagellin	
30	33	55.9	1095	1	W80359	Actin F-actin-combine	
31	33	55.9	954	1	W72752	DNA polymerase I epsilon	
32	33	55.9	464	1	W72907	Mycobacterium tuberculosis	
33	33	55.9	731	1	W87767	Human gelosolin, N-glycosylated	
34	33	55.9	688	1	W98709	H. pylori GHPO 758	
35	33	55.9	147	1	W98558	H. pylori GHPO 169	
36	33	55.9	224	1	Y06931	C. albicans antigen	
37	32	54.2	295	1	R27114	prad1, New cyclin	
38	32	54.2	295	1	R54044	bcl-1 protein, Nucleosome	
39	32	54.2	896	1	W08796	High osmolality G1	
40	32	54.2	459	1	W20736	H. pylori inner membrane	
41	32	54.2	660	1	W18570	Human cyclin D1-hu	
42	32	54.2	662	1	W18572	Human cyclin D1-hu	
43	32	54.2	605	1	W18573	Human cyclin D1-hu	

US/5/88902-A.
PD 04-AUG-1998.
PD 28-AUG-1996; 703947.
PF 17-JAN-1995; US-373957.
PR 17-JAN-1995; US-703947.
PR 28-AUG-1996; US-703947.
PA (UMOR) UNIV MISSOURI.
PI McIntosh MA, Wise KS;
DR WPI: 98-445004/38.
DR N-PSDB: V39862.
PT Vaccine against mycoplasma pneumonia in pigs - containing
PT Mycoplasma hyopneumoniae p65 surface antigen fusion protein
PS Claim 1: Fig 2; 29pp; English.
PS A vaccine has been developed for protecting pigs against mycoplasma
CC pneumonia caused by Mycoplasma hyopneumoniae. The vaccine contains a
CC immunogenic fusion protein comprising a first amino acid sequence
CC

CC to a second amino acid sequence, where the first sequence is the present
 CC sequence of 627 amino acids, which is a P65 surface antigen. The
 CC sequence given in V39862 represents the Mycoplasma hyopneumoniae
 CC surface lipoprotein P65 structural gene, which encodes the P65 surface
 CC antigen. The vaccine is used to induce an immune response in pigs
 CC against mycoplasma pneumonia caused by Mycoplasma hyopneumoniae.
 SQ Sequence 627 AA;

Query Match 89.8%; Score 53; DB 1; Length 627;
 Best Local Similarity 91.7%; Pred. NO. 0.13; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
 |:::|:::|
 DB 551 NLKPEQILTLG 562

RESULT 3

R57474
 ID R57474 standard; Protein; 3135 AA.
 AC R57474;
 DT 20-FEB-1995 (first entry)
 DE P. falciparum transmission blocking target antigen Pfs230.
 KW Protozoan; transmission blocking target antigen; Pfs230; malaria;
 OS vaccine.
 PN Plasmodium falciparum.
 PN W09417187-A.
 PD 04-AUG-1994.
 PF 18-JAN-1994; U00547.
 PR 29-JAN-1993; US-010409.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.
 PI Kaslow DC, Williamson KC;
 DR WPI; 94-264101/32.
 DR N-PSDB; Q67190.
 DE New Plasmodium falciparum transmission blocking target antigen -
 PT useful in antimalarial vaccines, also related DNA, expression
 PT vectors and transformed cells
 PS Claim 6; Page 24; 63pp; English.
 CC Pf230 protein is administered to humans to prevent transmission of
 CC malaria by inducing a transmission blocking immune response. It
 CC can also be used to raise antibodies and for T and B cell epitope
 CC mapping. Pfs230 induces a high and long-lasting antibody titer and
 CC can be produced in large amounts at low cost.
 SQ Sequence 3135 AA;

Query Match 61.0%; Score 36; DB 1; Length 3135;
 Best Local Similarity 45.5%; Pred. NO. 5.4e+02;
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
 |:::|:::|
 DB 2284 NLKPKDVIELM 2294

RESULT 4

W31365
 ID W31365 standard; Protein; 579 AA.
 AC W31365;
 DT 06-MAY-1998 (first entry)
 DE C16N for promoting neuron survival and type 1 collagen production.
 KW C16; C16N; neuron survival; type 1 collagen; calcium regulation;
 KW hypercalcaemia; hypertension; diabetes; arteriosclerosis; cancer;
 KW myocardial infarction; hydroxyapatite; osteoblast.
 OS Mus sp.
 PN W09740150-A1.
 PD 30-OCT-1997.
 PF 23-APR-1997; J01391.
 PR 10-FEB-1997; JP-041562.
 PR 23-APR-1996; JP-127954.
 PA (SUMU) SUMITOMO PHARM CO LTD.

PI Ishiduka Y, Mochizuki R;
 DR WPI; 97-535834/49.
 DR N-PSDB; V02312.
 PT Proteins C16 and C16N promote neuron survival and type 1 collagen
 PT production - for treatment of diseases involving collagen
 PT Claim 8; Page 58-60; 86pp; Japanese.
 CC The present sequence represents C16N which can: (a) induce
 CC differentiation into cells which can degrade hydroxyapatite; (b)
 CC maintain neuron survival; (c) inhibit osteoblast proliferation; and (d)
 CC promote type 1 collagen expression in osteoblasts. C16 and C16N are
 CC agents for the treatment of a broad range of diseases including
 CC hypercalcaemia, hypertension, diabetes, arteriosclerosis, myocardial
 CC infarction and terminal cancer. They may also be used as a screen for
 CC potential inhibitors of their activity for possible medicinal use.
 CC Transgenic animals containing DNA coding for the proteins can be used as
 CC model organisms and for the production of recombinant C16/C16N.
 SQ Sequence 579 AA;

Query Match 61.0%; Score 36; DB 1; Length 579;
 Best Local Similarity 66.7%; Pred. NO. 1e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 PEQILQLLG 12
 |:::|:::|
 DB 514 PQELQLLG 522

RESULT 5

W31366
 ID W31366 standard; Protein; 579 AA.
 AC W31366;
 DT 06-MAY-1998 (first entry)
 DE C16N for promoting neuron survival and type 1 collagen production.
 KW C16; C16N; neuron survival; type 1 collagen; calcium regulation;
 KW hypercalcaemia; hypertension; diabetes; arteriosclerosis; cancer;
 KW myocardial infarction; hydroxyapatite; osteoblast.
 OS Mus sp.
 PN W09740150-A1.
 PD 30-OCT-1997.
 PF 23-APR-1997; J01391.
 PR 10-FEB-1997; JP-041562.
 PR 23-APR-1996; JP-127954.
 PA (SUMU) SUMITOMO PHARM CO LTD.
 PI Ishiduka Y, Mochizuki R;
 DR WPI; 97-535834/49.
 DR N-PSDB; V02313.

PT Proteins C16 and C16N promote neuron survival and type 1 collagen
 PT production - for treatment of diseases involving collagen
 PT production, calcium regulation or neuron survival
 PS Claim 8; Page 63-66; 86pp; Japanese.
 CC The present sequence represents C16N which can: (a) induce
 CC differentiation into cells which can degrade hydroxyapatite; (b)
 CC maintain neuron survival; (c) inhibit osteoblast proliferation; and (d)
 CC promote type 1 collagen expression in osteoblasts. C16 and C16N are
 CC agents for the treatment of a broad range of diseases including
 CC hypercalcaemia, hypertension, diabetes, arteriosclerosis, myocardial
 CC infarction and terminal cancer. They may also be used as a screen for
 CC potential inhibitors of their activity for possible medicinal use.
 CC Transgenic animals containing DNA coding for the proteins can be used as
 CC model organisms and for the production of recombinant C16/C16N.
 SQ Sequence 579 AA;

Query Match 61.0%; Score 36; DB 1; Length 579;
 Best Local Similarity 66.7%; Pred. NO. 1e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 PEQILQLLG 12
 |:::|:::|
 DB 514 PQELQLLG 522

RESULT 6

R36601 ID R36601 standard; Protein; 567 AA.

AC R36601;

DT 27-AUG-1993 (first entry)

DE TGF-beta1 receptor type II (clone 3FF).

KW Transforming growth factor; receptor; antibody; diagnosis; binding.

OS Homo sapiens.

PN WO9309228-A.

PD 13-MAY-1993.

PF 30-OCT-1992; U093226.

PR 31-OCT-1991; US-786063.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Lin HY, Lodish HF, Wang X, Weinberg RA;

DR WPI: 93-167695/20.

DR N-PSDB; Q42383.

PT Transforming growth factor-beta receptor and corresp. DNA - encode type II and III receptors, for regulating effect of PT transforming growth factor in-vivo and in vitro

PS Claim 11; Fig 3; 56pp; English.

CC The full-length type II TGF-beta receptor cDNA clone 3FF was isolated from a human HepG2 cell cDNA library (full insert size 5 kb). The cDNA has an open reading frame encoding a 572 amino acid residue protein (sic - the fig. contg. the protein sequence comprises 567 amino acids).

CC The TGF-beta1 receptor and antibodies to it are used in a medicament to regulate the effect of TGF-beta, for assessing TGF-beta function in vitro and in vivo, and in diagnosis to detect abnormal binding.

CC The DNA may be used to identify equiv. TGF-beta receptor genes.

SQ Sequence 567 AA;

Query Match 59.3%; Score 35; DB 1; Length 567;
Best Local Similarity 72.7%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
III I I I I I

Db 298 NLKHENILQL 308

RESULT 7

R43342 ID R43342 standard; Protein; 1068 AA.

AC R43342;

DT 12-APR-1994 (first entry)

DE Human p110.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist; antagonist; cell proliferation; inhibition; prophylaxis; therapy; platelets; neutrophil activity; 3-phosphorylated phosphoinositides.

OS Human.

PH Key Location/Qualifiers

FT Domain 19..100

FT /note= "binds with p85 subunit"

PN WO9321328-A.

PD 28-OCT-1993.

PF 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

DR Parkerpj, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR N-PSDB; Q51156.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase activity, useful for controlling cell proliferation

PS Claim 24; Fig 16; 146pp; English.

CC Southern blot analysis was performed using a bovine cDNA probe contg. a fragment of a PI3-kinase-encoding sequence and human cDNA isolated from a cDNA library constructed from mRNA isolated from the human cell line KGla. Positive clones were sequenced to give a human PI3 kinase p110 sequence. This sequence has 95 percent homology with the bovine sequence. The domain encoding residues 19-100 of human p110 is sufficient to encode the kinase which will associate with the p85 kinase subunit. The gene may be used to provide a protein with PI3 kinase activity, and is useful for screening for (ant)agonists of PI3 kinase activity which could be useful for stimulation or inhibition of cell proliferation and hence prophylaxis or therapy. Platelet or neutrophil activity or blood glucose levels can be controlled using the kinase.

CC See also R43342 and R46552-3.

SQ Sequence 1068 AA;

Query Match 59.3%; Score 35; DB 1; Length 1068;
Best Local Similarity 60.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 LKPEQILQLL 11
:IIII :II

Db 593 IRPEQAMELL 602

RESULT 9

R88215 ID R88215 standard; Protein; 567 AA.

AC R88215;

DT 14-JUN-1996 (first entry)

DE Human TGF-beta type II receptor.

CC with the bovine sequence. The domain contg. residues 19-100 of human p110 is sufficient to associate with the p85 kinase subunit. The protein with PI3 kinase activity is useful for screening for (ant)agonists of PI3 kinase activity which could be useful for stimulation or inhibition of cell proliferation and hence prophylaxis or therapy. Platelet or neutrophil activity or blood glucose levels can be controlled using the kinase.

CC See also R43341 and R46552-3.

SQ Sequence 1068 AA;

Query Match 59.3%; Score 35; DB 1; Length 1068;
Best Local Similarity 60.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 LKPEQILQLL 11
:IIII :II

Db 593 IRPEQAMELL 602

RESULT 8

R43341 ID R43341 standard; Protein; 1068 AA.

AC R43341;

DT 12-APR-1994 (first entry)

DE p110.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist; antagonist; cell proliferation; inhibition; prophylaxis; therapy; platelets; neutrophil activity; 3-phosphorylated phosphoinositides.

PN WO9321328-A.

PD 28-OCT-1993.

PF 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

DR Parkerpj, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR N-PSDB; Q51155.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase activity, useful for controlling cell proliferation

PS Claim 4; Fig 9; 146pp; English.

CC Southern blot analysis was performed using a bovine cDNA probe contg. a fragment of a PI3-kinase-encoding sequence and human cDNA isolated from a cDNA library constructed from mRNA isolated from the human cell line KGla. Positive clones were sequenced to give a human PI3 kinase p110 sequence. This sequence has 95 percent homology with the bovine sequence. The domain encoding residues 19-100 of human p110 is sufficient to encode the kinase which will associate with the p85 kinase subunit. The gene may be used to provide a protein with PI3 kinase activity, and is useful for screening for (ant)agonists of PI3 kinase activity which could be useful for stimulation or inhibition of cell proliferation and hence prophylaxis or therapy. Platelet or neutrophil activity or blood glucose levels can be controlled using the kinase.

CC See also R43342 and R46552-3.

SQ Sequence 1068 AA;

KW TGF-beta; transforming growth factor; cytokine; fetuin; TRH1;
 KW fibrosis; tumour; bone formation; autoimmune disease; wound healing;
 KW binding; transplacental rejection; hypotension.
 OS Homo sapiens.
 PN W09530900-A1.
 PD 16-NOV-1995.
 PF 04-MAY-1995; CA0290.
 PR 04-MAY-1994; US-237715.
 PA (MOUN) MOUNT SINAI HOSPITAL CORP.
 PI Demetriou M, Dennis JW;
 DR WPI; 95-404212/51.
 PT Modulators of cytokine(s) of the TGF-beta super-family - obtd. using
 PT cpds. which contain a TGF-beta receptor II homology I domain useful
 PT for e.g. modulating bone formation, treating fibrosis, auto-immune
 PT disorders, etc
 PS Disclosure: Page 48-50; 79pp; English.
 CC R8215 is the human transforming growth factor (TGF) beta type II
 CC receptor. This protein contains a TRH1 domain which can be used
 CC in a method for assaying for the presence of a substance that
 CC modulates cytokines of the TGF-beta superfamily. TGF-beta binding
 CC compounds can be used for treating conditions requiring modulation of
 CC TGF-beta cytokines. They can also be used for modulating bone formation
 CC and for treating fibrosis, autoimmune disorders, proliferative disorders,
 CC tumours, viral infections, transplant rejection or hypotension. A further
 CC use of TGF-beta binding compounds is to promote wound healing.
 SQ Sequence 567 AA;

Query Match 59.3%; Score 35; DB 1; Length 567;
 Best Local Similarity 72.7%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
 ||| | ||| |
 Db 298 NLKHENILQL 308

RESULT 10

W20733
 ID W20733 standard; protein; 432 AA.
 AC W20733; 1997 (first entry)
 DE H. pylori cell envelope protein, 06cp1172orf15.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI; 97-052306/05.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 56; Page 1148-1149; 1481pp; English.
 CC The present sequence is a Helicobacter pylori cell envelope protein.
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide

CC production, e.g. in E. coli hosts.
 SQ Sequence 432 AA;

Query Match 59.3%; Score 35; DB 1; Length 432;
 Best Local Similarity 50.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 12
 ||| | ||| |
 Db 405 NLKADDEVLLILG 416

RESULT 11

W20358
 ID W20358 standard; protein; 100 AA.
 AC W20358;
 DT 11-JUL-1997 (first entry)
 DE H. pylori cell envelope protein 26366312.aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; U09122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI; 97-052306/05.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 56; Page 549; 1481pp; English.
 CC This sequence is a H. pylori cell envelope protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds.
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SQ Sequence 100 AA;

Query Match 59.3%; Score 35; DB 1; Length 100;
 Best Local Similarity 50.0%; Pred. No. 27;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 12
 ||| | ||| |
 Db 57 NLKADDEVLLILG 68

RESULT 12

P94623
 ID P94623 standard; protein; 304 AA.
 AC P94623;
 DT 22-JUN-1990 (first entry).
 DE Portion of clone P4A encoding flagellin gene of Bacillus subtilis.
 KW Flagellin; fusion protein; heterologous protein; ds.
 OS Bacillus subtilis.
 PN US4801536-A.
 PD 31-JAN-1989.
 PF 2-JUN-1987; 57881.

PR 11-OCT-1985; US-786749.
 PR 10-OCT-1986; WO-U02168.
 PR 02-JUN-1987; US-57881.
 PA (GENE-) Genetics Inst Inc.
 PI Stahl ML, Lavallie ER;
 DR WPI: 89-053638/07.
 DR N-PSDB; N91115.

PT Prodn. of heterologous protein in bacterial host cells -
 PT exported into the culture medium as fusion protein with
 PT flagellin protein native to host.

PS Disclosure; 4pp; English.
 CC By ligating the flagellin coding sequence to a protein required to be
 CC secreted into the medium, the product can be collected. If a linker
 CC encoding a selectively cleavable peptide is used in ligation, flagellin
 CC portion may be cleaved from the product.

SQ Sequence 304 AA;

Query Match 57.6%; Score 34; DB 1; Length 304;
 Best Local Similarity 54.5%; Pred. No. 1.2e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
 | : : : :
 DB 293 NQPPNVQLLL 303

RESULT 13

ID P94263 standard; protein; 492 AA.

AC P94263; (first entry)
 DT 19-JUN-1990
 DE Bovine adrenal gland adrenodoxin reduction enzyme;
 KW Bovine adrenal gland adrenodoxin reduction enzyme;
 KW electron transport system.

OS Bos taurus;

PN J01010989-A.

PD 13-JAN-1989.

PF 30-JUN-1987; 165084.

PR 30-JUN-1987; JP-165084.

PA (KATA/) Katanaka Y.

DR WPI: 89-058104/08.

DR N-PSDB; N91045.

PT Gene encoding bovine adrenal gland adrenodoxin reduction enzyme - used in
 PT prodn. of enzymes composing electron transport system of bovine adrenal
 PT gland mitochondria.

PS Disclosure; 2; 7pp; Japanese.

CC The adrenodoxin reduction enzyme produces all the enzymes composing the
 CC electron transport system of bovine adrenal gland mitochondria.

SQ Sequence 492 AA;

Query Match 57.6%; Score 34; DB 1; Length 492;
 Best Local Similarity 54.5%; Pred. No. 1.9e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 LKPEQILQLLG 12
 | : : : :
 DB 481 LDPQEMRLLLG 491

RESULT 14

ID P70637 standard; Protein; 304 AA.

AC P70637;
 DT 18-APR-1991 (first entry)
 DE Sequence encoded by flagellin gene of B. subtilis in clone p4A.
 KW Heterologous protein prodn; flagellin.

OS B. subtilis.

PN W08702385-A.

PD 23-APR-1987.

PF 10-OCT-1986; U02168.

PR 11-OCT-1985; US-786749.

PR 02-JUN-1987; US-057881.
 PA (GENE-) GENETICS INST.
 PI Stahl ML, Lavallie E;
 DR WPI: 87-122742/17.
 DR N-PSDB; N70984.

PT Heterologous protein prodn. in bacterial host cell - using
 PT nucleotide sequence encoding at least N-terminal portion of
 PT flagellin protein

PS Disclosure; Table 1, p10A; 42pp; English.

CC The DNA sequence encoding the N-terminal portion of flagellin, eg
 CC a portion of the hag gene of B. subtilis, is operatively linked to
 CC an expression control sequence. The flagellin gene is linked via its
 CC 3' terminus to the 5' terminus of a heterologous protein via a
 CC linking nucleotide sequence which encodes a selectively cleavable
 CC polypeptide such that the exported protein contains a selectively
 CC cleavable site. Proteins which may be produced include hormones.

SQ Sequence 304 AA;

Query Match 57.6%; Score 34; DB 1; Length 304;
 Best Local Similarity 54.5%; Pred. No. 1.2e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
 | : : : :
 DB 293 NQPPNVQLLL 303

RESULT 15

ID W13135 standard; Protein; 794 AA.

AC W13135;

DT 14-MAY-1997 (first entry)

DE Putative human cadherin-12.

KW Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;

KW brain; human; antibody; purification; determination;

KW tissue expression; binding antagonist; calcium ion; putative.

OS Homo sapiens.

PN US5597725-A.

PD 28-JAN-1997.

PF 17-APR-1992; 872643.

PR 17-APR-1992; US-872643.

PR 19-APR-1993; US-049460.

PR 26-JAN-1994; US-188228.

PA (DOHE-) DOHENY EYE INST.

PI Suzuki S;

DR WPI: 97-108328/10.

DR N-PSDB; T61926.

PT Antibodies to cadherin proteins - useful as cadherin antagonists,

PT etc.

PS Example 2; Columns 101-106; 59pp; English.

CC The present sequence is a putative human cadherin-12, which

CC is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA

CC was isolated from a foetal brain cDNA library, using probes based

CC on homologous rat cadherin cDNA.

CC Antibodies or fragments that specifically bind the human cadherin

CC can be used to purify the cadherin, determine its tissue expression

CC and antagonise its ligand/antiligand binding activities.

SQ Sequence 794 AA;

Query Match 57.6%; Score 34; DB 1; Length 794;
 Best Local Similarity 54.5%; Pred. No. 3.1e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
 | : : : :
 DB 500 NAKPGQIIQIV 510

Search completed: November 13, 1999, 19:00:03
 Job time: 142 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:35 ; Search time 64.87 Seconds
(without alignments)
2.114 Million cell updates/sec

Title: US-08-913-430-14
Perfect score: 59
Sequence: 1 NUKPEOILOLLG 12

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued_Patents_AA.*
1: /cgn2_5/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/PCTUS9_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	89.8	627	2	US-08-703-947-2
2	36	61.0	3135	1	US-08-323-170B-2
3	35	59.3	567	1	US-08-361-873A-2
4	35	59.3	567	2	US-08-483-926A-1
5	35	59.3	1080	2	US-08-162-081B-36
6	35	59.3	1069	2	US-08-162-081B-37
7	35	59.3	567	2	US-08-854-768-1
8	35	59.3	565	2	US-08-357-533A-9
9	35	59.3	1080	2	US-08-780-872-36
10	35	59.3	1069	2	US-08-780-872-37
11	35	59.3	565	2	US-08-459-009-9
12	35	59.3	567	2	US-08-445-520B-9
13	35	59.3	567	3	PCT-US92-09326-4
14	34	57.6	794	1	US-08-188-228-60
15	34	57.6	794	1	US-08-332-643-54
16	34	57.6	794	1	US-08-332-638-60
17	33	55.9	782	1	US-07-725-083-2
18	33	55.9	877	1	US-08-208-036-14
19	33	55.9	171	1	US-08-193-977-5
20	33	55.9	877	2	US-08-428-823-14
21	33	55.9	693	2	US-08-463-620-11
22	33	55.9	876	2	US-08-510-215A-2
23	33	55.9	876	2	US-08-633-476-2
24	33	55.9	1572	2	US-08-290-731C-5
25	33	55.9	430	2	US-08-290-731C-9
26	33	55.9	101	2	US-08-580-988A-22
27	33	55.9	148	2	US-08-460-694-6
28	33	55.9	819	2	US-08-464-517-7
29	33	55.9	106	2	US-08-464-517-26
30	33	55.9	876	2	US-08-436-664-20
31	33	55.9	876	2	US-08-436-664-32
32	33	55.9	876	2	US-08-436-664-34
33	33	55.9	819	3	PCT-US93-05000-7
34	33	55.9	116	3	PCT-US93-05000-26
35	33	55.9	693	3	PCT-US95-03934A-11
36	33	55.9	876	3	PCT-US95-04080-20
37	33	55.9	876	3	PCT-US95-04080-32
38	33	55.9	876	3	PCT-US95-04080-34
39	32	54.2	295	1	US-07-947-120-8

Sequence 7, Appli
Sequence 8, Appli
Sequence 2, Appli
Sequence 3, Appli
Sequence 5, Appli
Sequence 7, Appli

40 32 54.2 173 1 US-08-193-977-7
41 32 54.2 295 1 US-08-472-893A-8
42 32 54.2 660 2 US-08-770-761A-2
43 32 54.2 618 2 US-08-770-761A-3
44 32 54.2 662 2 US-08-770-761A-5
45 32 54.2 705 2 US-08-770-761A-7

ALIGNMENTS

RESULT 1
US-08-703-947-2
; Sequence 2, Application US/08703947
; Patent No. 5788962
; GENERAL INFORMATION:
; APPLICANT: Wise, Kim S.
; APPLICANT: McIntosh, Mark A.
; TITLE OF INVENTION: DNA Sequences Coding for Mycoplasma
; TITLE OF INVENTION: Hypopneumoniae Surface Antigens,
; TITLE OF INVENTION: Corresponding Proteins and Use in
; TITLE OF INVENTION: Vaccines and Diagnostic Procedures
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace J. Fishel
; STREET: 929 Fee Fee Road, Suite 100
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63043
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
; COMPUTER: Hewlett-Packard Vectra
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,947
; FILING DATE: 28-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/373,957
; FILING DATE: January 17, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fishel, Grace J.
; REGISTRATION NUMBER: 25864
; REFERENCE/DOCKET NUMBER: UVM 8141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 878-0440
; TELEFAX: (314) 275-7693
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 amino acid residues
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: predicted amino acid sequence of complete
; DESCRIPTION: 627 residues of the p65 lipoprotein, derived
; DESCRIPTION: from the nucleic acid sequence
; HYPOTHETICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: whole polypeptide
; ORIGINAL SOURCE:
; ORGANISM: Mycoplasma hyopneumoniae
; STRAIN: J
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE: unicellular bacterium
; CELL LINE:
; ORGANELLE:


```

; IMMEDIATE SOURCE:
; LIBRARY: Genomic in Charon 4A, GEM12
; CLONE: MhpJ25, MhpJ35, MhpJG35, pJ25.1, pJ25.14,
; CLONE: pJG35.1, pJG35.12, pJG35.13, pJG35.14
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: single chromosome
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: 627 amino acid sequence representing
; NAME/KEY: complete sequence (including signal
; NAME/KEY: sequence) of surface lipoprotein P65
; LOCATION: entire derived coded sequence
; IDENTIFICATION METHOD: clone identified by immunodetection of
; IDENTIFICATION METHOD: protein product with antiserum specific for
; IDENTIFICATION METHOD: P65; residue sequence deduced from nucleic
; IDENTIFICATION METHOD: acid sequence
; OTHER INFORMATION: immunogenic surface lipoprotein of no known
; OTHER INFORMATION: function; C-terminus exposed on external
; OTHER INFORMATION: surface of cell; N-terminal signal sequence
; OTHER INFORMATION: (first 29 amino acid residues) cleaved during
; OTHER INFORMATION: lipid modification process
; PUBLICATION INFORMATION:
; AUTHORS: Mary F. Kim, Manijeh B. Heidari, Susan J.
; AUTHORS: Stull, Mark A. McIntosh, and Kim S. Wise
; TITLE: Identification and Mapping of an
; TITLE: Immunogenic Region of Mycoplasma
; TITLE: hyopneumoniae p65 Surface Lipoprotein
; TITLE: Expressed in Escherichia coli from a Cloned
; TITLE: Genomic Fragment
; JOURNAL: Infection and Immunity
; VOLUME: 58
; ISSUE: 8
; PAGES: 2637-2643
; DATE: August 1990
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO: From 1 to 627
; US-08-703-947-2

Query Match 89.8%; Score 53; DB 2; Length 627;
Best Local Similarity 91.7%; Pred. No. 0.053;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
Db 551 NLKPEQILTLG 562

RESULT 2
US-08-323-170B-2
; Sequence 2, Application US/08323170B
; Patent No. 5733772
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170B
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 015280-113100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-323-170B-2

Query Match 61.0%; Score 36; DB 1; Length 3135;
Best Local Similarity 45.5%; Pred. No. 2.9e+02;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
Db 2284 NLKPKDVIEM 2294

RESULT 3
US-08-361-873A-2
; Sequence 2, Application US/08361873A
; Patent No. 5693607
; GENERAL INFORMATION:
; APPLICANT: SEGARINI, PATRICIA R.
; APPLICANT: DASCH, JAMES R.
; APPLICANT: OLSEN, DAVID R.
; APPLICANT: CARRILLO, PEDRO A.
; TITLE OF INVENTION: USES OF TGB-beta RECEPTOR FRAGMENT AS A
; TITLE OF INVENTION: THERAPEUTIC AGENT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,873A
; FILING DATE: 22-DEC-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: PARK, FREDDIE K.
; REGISTRATION NUMBER: 35,636
; REFERENCE/DOCKET NUMBER: 22095-20261.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 567 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
; US-08-361-873A-2

Query Match 59.3%; Score 35; DB 1; Length 567;
Best Local Similarity 72.7%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||| | ||| |
Db 298 NLKHENILQFL 308

RESULT 4
US-08-483-926A-1
; Sequence 1, Application US/08483926A
; Patent No. 5821227
; GENERAL INFORMATION:
; APPLICANT: Dennis, James W.
; TITLE OF INVENTION: MODULATORS OF CYTOKINES OF THE TGF BETA
; TITLE OF INVENTION: SUPERFAMILY AND METHODS FOR ASSAYING FOR SAME
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,926A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurdydyk, Linda M.
; REGISTRATION NUMBER: 34,971
; REFERENCE/DOCKET NUMBER: 3153-155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 567 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; LIBRARY: Lambda zapII
US-08-483-926A-1

Query Match 59.3%; Score 35; DB 2; Length 567;
Best Local Similarity 72.7%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||| | ||| |
Db 298 NLKHENILQFL 308

RESULT 5
US-08-162-081B-36
; Sequence 36, Application US/08162081B
; Patent No. 5824492
; GENERAL INFORMATION:

; APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu
; APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter
; APPLICANT: Joseph; Otsu, Masayuki; Panayotou, George; Volinia,
; APPLICANT: Stefano; Gout, Ivan Tarasovitch
; TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,
; TITLE OF INVENTION: THEIR PREPARATION AND USE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,081B
; FILING DATE: February 7, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00761
; FILING DATE: 13 April 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: LUD 5256
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1080 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-162-081B-36

Query Match 59.3%; Score 35; DB 2; Length 1080;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 LKPEQILQLL 11
: ||| : |||
Db 593 IKPEQAMELL 602

RESULT 6
US-08-162-081B-37
; Sequence 37, Application US/08162081B
; Patent No. 5824492
; GENERAL INFORMATION:
; APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu
; APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter
; APPLICANT: Joseph; Otsu, Masayuki; Panayotou, George; Volinia,
; APPLICANT: Stefano; Gout, Ivan Tarasovitch
; TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,
; TITLE OF INVENTION: THEIR PREPARATION AND USE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162.081B
FILING DATE: February 7, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCY/GB93/00761
FILING DATE: 13 April 1993
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: LUD 5256
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 638-3884
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1069 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-162-081B-37

Query Match 59.3%; Score 35; DB 2; Length 1069;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 LKPEQILQLL 11
:||||:|
DB 593 IKPEQAMELL 602

RESULT 7
US-08-854-768-1
Sequence 1, Application US/08854768
Patent No. 5830671
GENERAL INFORMATION:
APPLICANT: Dennis, James W.
APPLICANT: Demetriou, Michael
TITLE OF INVENTION: MODULATORS OF CYTOKINES OF THE TGF
TITLE OF INVENTION: SUPERFAMILY AND METHODS FOR ASSAYING FOR SAME
NUMBER OF SEQUENCES: 7
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,768
FILING DATE: 12-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/237,715
FILING DATE: 04-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kurdýdyk, Linda M.
REGISTRATION NUMBER: 34,971
REFERENCE/DOCKET NUMBER: 3153-79
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 567 amino acids

TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
LIBRARY: Lambda zapII
US-08-854-768-1

Query Match 59.3%; Score 35; DB 2; Length 567;
Best Local Similarity 72.7%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
:||||:|
DB 298 NLKHENILOFL 308

RESULT 8
US-08-357-533A-9
Sequence 9, Application US/08357533A
Patent No. 5831050
GENERAL INFORMATION:
APPLICANT: JIN, DONALD F
APPLICANT: OPPERMAN, HERMANN
APPLICANT: KUBERASAMPATH, THANGAVEL
APPLICANT: SMART, JOHN E
TITLE OF INVENTION: NOVEL MORPHOGEN CELL SURFACE RECEPTOR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES,
ADDRESSEE: INC
STREET: 45 SOUTH STREET
CITY: HOPKINTON
STATE: MA
COUNTRY: USA
ZIP: 01748
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/357,533A
FILING DATE: 16-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KELLY, ROBIN D
REGISTRATION NUMBER: 34,637
REFERENCE/DOCKET NUMBER: CRP-073FW
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)-435-9001
TELEFAX: (508)-435-0992
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..565
OTHER INFORMATION: /note= "TGF-B TYPE II RECEPTOR"
US-08-357-533A-9

Query Match 59.3%; Score 35; DB 2; Length 565;
Best Local Similarity 72.7%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11

APPLICATION NUMBER: US/08/459,009

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,533
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: KELLY, ROBIN D
; REGISTRATION NUMBER: 34,637
; REFERENCE/DOCKET NUMBER: CRP-073FW
; TELEPHONE: (508)-435-9001
; TELEFAX: (508)-435-0992
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..565
; OTHER INFORMATION: /note= "TGF-B TYPE II RECEPTOR"
US-08-459-009-9

Query Match 59.3%; Score 35; DB 2; Length 565;
Best Local Similarity 72.7%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||| | ||| |
DB 298 NLKHENILQFL 308

RESULT 12
US-08-445-520B-9
; Sequence 9, Application US/08445520B
; Patent No. 5866323
; GENERAL INFORMATION:
; APPLICANT: Markowitz, Sanford D.
; APPLICANT: Brattain, Michael G.
; APPLICANT: Willson, James K.V.
; TITLE OF INVENTION: CANCER DIAGNOSIS, PROGNOSIS AND
; TITLE OF INVENTION: THERAPY BASED ON MUTATION OF RECEPTOR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & BOTTS, L.L.P.
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004-2400
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,520B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/417,867
; FILING DATE: 07-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Posorske, Laurence H
; REGISTRATION NUMBER: 34,698
; REFERENCE/DOCKET NUMBER: 062361-0101
; TELEPHONE: 202-639-7700
; TELEFAX: 202-639-7890
; TELEX:

; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 567 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-445-520B-9

Query Match 59.3%; Score 35; DB 2; Length 567;
Best Local Similarity 72.7%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||| | ||| |
DB 298 NLKHENILQFL 308

RESULT 13
PCT-US92-09326-4
; Sequence 4, Application PC/TUS9209326
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; TITLE OF INVENTION: TGF-BETA TYPE RECEPTOR CDNAS ENCODED
; TITLE OF INVENTION: PRODUCTS AND USES THEREFOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09326
; FILING DATE: 19921030
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/786,063
; FILING DATE: 31-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI91-09 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 567 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US92-09326-4

Query Match 59.3%; Score 35; DB 3; Length 567;
Best Local Similarity 72.7%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||| | ||| |
DB 298 NLKHENILQFL 308

RESULT 14
US-08-188-228-60
; Sequence 60, Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,228
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 794 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-228-60

Query Match 57.6%; Score 34; DB 1; Length 794;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
Db 500 NAKPGQIIQIV 510

RESULT 15
US-08-332-643-54
; Sequence 54, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA

; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 794 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-332-643-54

Query Match 57.6%; Score 34; DB 1; Length 794;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
Db 500 NAKPGQIIQIV 510

Search completed: November 13, 1999, 10:56:36
Job time: 1365 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 05:08:42 ; Search time 251.81 Seconds
(without alignments)
3.017 Million cell updates/sec

Title: US-08-913-430-14
Perfect score: 59
Sequence: 1 NLKPEQIQLLG 12

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA.*
1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	100.0	12	13	US-08-913-430-14
2	53	89.8	627	6	US-08-373-957-2
3	41	69.5	1203	14	US-09-060-939A-2
4	41	69.5	1203	16	US-09-207-857-2
5	41	69.5	1203	16	US-09-293-505-2
6	38	64.4	1182	14	US-09-060-939A-7
7	38	64.4	1013	16	US-09-248-796-18605
8	38	64.4	1182	16	US-09-293-505-7
9	38	64.4	961	17	US-09-328-352-7358
10	38	64.4	1013	19	US-60-096-409-18605
11	37	62.7	291	17	US-09-328-352-4571
12	37	62.7	182	23	US-09-417-507-30182
13	36	61.0	3135	1	PCT-US94-00547-2
14	36	61.0	3135	6	US-08-323-170-2
15	36	61.0	3135	3	US-08-323-170A-2
16	36	61.0	3135	13	US-08-954-441-2
17	36	61.0	579	15	US-09-171-710-4
18	36	61.0	579	15	US-09-171-710-6
19	36	61.0	1201	17	US-09-328-352-5890

20	35	59.3	566	3	US-08-067-462-3	Sequence 3, Appli
21	35	59.3	565	3	US-08-073-199-9	Sequence 9, Appli
22	35	59.3	325	4	US-08-158-735A-13	Sequence 13, Appli
23	35	59.3	567	6	US-08-311-703A-8	Sequence 8, Appli
24	35	59.3	1068	6	US-08-381-840-2	Sequence 2, Appli
25	35	59.3	1068	6	US-08-390-874C-11	Sequence 11, Appli
26	35	59.3	567	7	US-08-445-520A-11	Sequence 11, Appli
27	35	59.3	567	7	US-08-446-936A-8	Sequence 8, Appli
28	35	59.3	567	7	US-08-446-937B-8	Sequence 8, Appli
29	35	59.3	567	7	US-08-446-938B-8	Sequence 8, Appli
30	35	59.3	567	7	US-08-446-939B-8	Sequence 8, Appli
31	35	59.3	567	7	US-08-451-946B-8	Sequence 8, Appli
32	35	59.3	565	7	US-08-459-951-9	Sequence 9, Appli
33	35	59.3	567	8	US-08-483-926-1	Sequence 1, Appli
34	35	59.3	100	9	US-08-487-032A-657	Sequence 657, App
35	35	59.3	100	9	US-08-561-469A-657	Sequence 657, App
36	35	59.3	1259	10	US-08-695-177-13	Sequence 13, Appli
37	35	59.3	567	11	US-08-737-045-1	Sequence 1, Appli
38	35	59.3	567	11	US-08-737-045-1	Sequence 1, Appli
39	35	59.3	432	11	US-08-761-184-1158	Sequence 1158, Ap
40	35	59.3	447	11	US-08-761-184-1590	Sequence 1590, Ap
41	35	59.3	432	12	US-08-821-931-1158	Sequence 1158, Ap
42	35	59.3	447	12	US-08-821-931-1590	Sequence 1590, Ap
43	35	59.3	1068	13	US-08-908-453-10	Sequence 10, Appli
44	35	59.3	1080	14	US-09-085-957-36	Sequence 36, Appli
45	35	59.3	1069	14	US-09-085-957-37	Sequence 37, Appli

ALIGNMENTS

RESULT 1
US-08-913-430-14
Sequence 14, Application US/08913430B
GENERAL INFORMATION:
APPLICANT: Walker, John
APPLICANT: Lee, Rogan
TITLE OF INVENTION: Antigen Composition Against Mycoplasma
FILE REFERENCE: U-011415-0
CURRENT APPLICATION NUMBER: US/08/913,430B
CURRENT FILING DATE: 1997-12-09
EARLIER APPLICATION NUMBER: PCT/AU96/00149
EARLIER FILING DATE: 1996-03-15
EARLIER APPLICATION NUMBER: PN 1789
EARLIER FILING DATE: 1995-03-16
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0 - beta
SEQ ID NO 14
LENGTH: 12
TYPE: PRT
ORGANISM: Mycoplasma hyopneumoniae
US-08-913-430-14

Query Match 100.0%; Score 59; DB 13; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00039;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLKPEQIQLLG 12
|||||
DB 1 NLKPEQIQLLG 12

RESULT 2
US-08-373-957-2
Sequence 2, Application US/08373957
GENERAL INFORMATION:
APPLICANT: Wise, Kim S.
APPLICANT: McIntosh, Mark A.
TITLE OF INVENTION: DNA Sequences Coding for Mycoplasma
TITLE OF INVENTION: Hypopneumoniae Surface Antigens,
TITLE OF INVENTION: Corresponding Proteins and Use in Vaccines

TITLE OF INVENTION: and Diagnostic Procedures
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grace J. Fishel
STREET: 929 Fee Fee Road, Suite 100
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63043
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
COMPUTER: Hewlett-Packard Vectra
OPERATING SYSTEM: MS-DOS Version 3.3
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/373,957
FILING DATE: January 17, 1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fishel, Grace J.
REGISTRATION NUMBER: 25864
REFERENCE/DOCKET NUMBER: UVM 8141
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 878-0440
TELEFAX: (314) 275-7693
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 627 amino acid residues
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: predicted amino acid sequence of complete
DESCRIPTION: 627 residues of the P65 lipoprotein, derived
DESCRIPTION: from the nucleic acid sequence
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: whole polypeptide
ORIGINAL SOURCE:
ORGANISM: Mycoplasma hyopneumoniae
STRAIN: J
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE: unicellular bacterium
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY: Genomic in Charon 4A, GEM12
CLONE: MhpJ25, MhpJ35, MhpJG35, pJ25,
CLONE: pJ25.1, pJ25.14, pJ25.14, pJ25.14, pJ25.14,
CLONE: pJ25.13, pJ25.13
POSITION IN GENOME:
CHROMOSOME/SEGMENT: single chromosome
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: 627 amino acid sequence representing
NAME/KEY: complete sequence (including signal
NAME/KEY: sequence) of surface lipoprotein P65
LOCATION: entire derived coded sequence
IDENTIFICATION METHOD: clone identified by immunodetection of
IDENTIFICATION METHOD: protein product with antiserum specific for
IDENTIFICATION METHOD: P65; residue sequence deduced from nucleic
IDENTIFICATION METHOD: acid sequence
OTHER INFORMATION: immunogenic surface lipoprotein of no known
OTHER INFORMATION: function; C-terminus exposed on external
OTHER INFORMATION: surface of cell; N-terminal signal sequence

OTHER INFORMATION: (first 29 amino acid residues) cleaved during
OTHER INFORMATION: lipid modification process
PUBLICATION INFORMATION:
AUTHORS: Mary F. Kim, Manijeh B. Heidari, Susan J.
AUTHORS: Stull, Mark A. McIntosh, and Kim S. Wise
TITLE: Identification and Mapping of an
TITLE: Immunogenic Region of Mycoplasma
TITLE: hyopneumoniae p65 Surface Lipoprotein
TITLE: Expressed in Escherichia coli from a Cloned
TITLE: Genomic Fragment
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 8
PAGES: 2637-2643
DATE: August 1990
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: From 1 to 627
US-08-373-957-2

Query Match 89.8%; Score 53; DB 6; Length 627;
Best Local Similarity 91.7%; Pred. No. 0.39;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12

|||||

DB 551 NLKPEQILQLLG 562

RESULT 3

US-09-060-939A-2
Sequence 2, Application US/09060939A
GENERAL INFORMATION:
APPLICANT: Frederic de Sauvage, David A. Carpenter
TITLE OF INVENTION: Patched-2
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/060,939A
FILING DATE: 15-Apr-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1405
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1203 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-09-060-939A-2

Query Match 69.5%; Score 41; DB 14; Length 1203;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
 || |||:|:|
 Db 215 NLDPEQLLEELG 226

RESULT 4

US-09-207-857-2
 ; Sequence 2, Application US/09207857
 ; GENERAL INFORMATION:
 ; APPLICANT: Bumcrot, David A.
 ; TITLE OF INVENTION: HUMAN PATCHED GENES AND PROTEINS, AND USES RELATED
 ; FILE REFERENCE: ONV-05001
 ; CURRENT APPLICATION NUMBER: US/09/207,857
 ; CURRENT FILING DATE: 1998-12-08
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 1203
 ; TYPE: PRT
 ; ORGANISM: human
 US-09-207-857-2

Query Match 69.5%; Score 41; DB 16; Length 1203;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
 || |||:|:|
 Db 215 NLDPEQLLEELG 226

RESULT 5

US-09-293-505-2
 ; Sequence 2, Application US/09293505
 ; GENERAL INFORMATION:
 ; APPLICANT: de Sauvage, Frederic
 ; APPLICANT: Carpenter, David A.
 ; TITLE OF INVENTION: Patched-2
 ; FILE REFERENCE: P1405R1
 ; CURRENT APPLICATION NUMBER: US/09/293,505
 ; CURRENT FILING DATE: 1999-04-15
 ; EARLIER APPLICATION NUMBER: US 60/081,884
 ; EARLIER FILING DATE: 1998-04-15
 ; NUMBER OF SEQ ID NOS: 32
 ; SEQ ID NO 2
 ; LENGTH: 1203
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-293-505-2

Query Match 69.5%; Score 41; DB 16; Length 1203;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
 || |||:|:|
 Db 215 NLDPEQLLEELG 226

RESULT 6

US-09-060-939A-7
 ; Sequence 7, Application US/09060939A
 ; GENERAL INFORMATION:
 ; APPLICANT: Frederic de Sauvage, David A. Carpenter
 ; TITLE OF INVENTION: Patched-2
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 1 DNA Way
 ; CITY: South San Francisco

; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/060,939A
 ; FILING DATE: 15-Apr-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Svoboda, Craig G.
 ; REGISTRATION NUMBER: 39,044
 ; REFERENCE/DOCKET NUMBER: P1405
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650/225-1489
 ; TELEFAX: 650/952-9881
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1182 amino acids
 ; TYPE: Amino Acid
 ; TOPOLOGY: Linear
 US-09-060-939A-7

Query Match 64.4%; Score 38; DB 14; Length 1182;
 Best Local Similarity 58.3%; Pred. No. 3.5e+02;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
 || |||:|:|
 Db 215 NLDPEQLLEELG 226

RESULT 7

US-09-248-796-18605.
 ; Sequence 18605, Application US/09248796
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstock et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796
 ; CURRENT FILING DATE: 1999-02-12
 ; NUMBER OF SEQ ID NOS: 28206
 ; SEQ ID NO 18605
 ; LENGTH: 1013
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 US-09-248-796-18605

Query Match 64.4%; Score 38; DB 16; Length 1013;
 Best Local Similarity 63.6%; Pred. No. 3e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LKPEQILQLLG 12
 | | | : | : |
 Db 132 LSPQDILQLLG 142

RESULT 8

US-09-293-505-7
 ; Sequence 7, Application US/09293505
 ; GENERAL INFORMATION:
 ; APPLICANT: de Sauvage, Frederic
 ; APPLICANT: Carpenter, David A.
 ; TITLE OF INVENTION: Patched-2
 ; FILE REFERENCE: P1405R1
 ; CURRENT APPLICATION NUMBER: US/09/293,505
 ; CURRENT FILING DATE: 1999-04-15

; EARLIER APPLICATION NUMBER: US 60/081,884
; EARLIER FILING DATE: 1998-04-15
; NUMBER OF SEQ ID NOS: 32
; SEQ ID NO 7
; LENGTH: 1182
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-293-505-7

Query Match 64.4%; Score 38; DB 16; Length 1182;
Best Local Similarity 58.3%; Pred. No. 3.5e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
|||:|:|
Db 215 NLDPOQLLELG 226

RESULT 9
US-09-328-352-7358
; Sequence 7358, Application US/09328352
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7358
; LENGTH: 961
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7358

Query Match 64.4%; Score 38; DB 17; Length 961;
Best Local Similarity 77.8%; Pred. No. 2.8e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 KPEQILQLL 11
||||:|
Db 139 KPEQIIELL 147

RESULT 10
US-09-096-409-18605
; Sequence 18605, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 18605
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-18605

Query Match 64.4%; Score 38; DB 19; Length 1013;
Best Local Similarity 63.8%; Pred. No. 3e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LKPEQILQLLG 12
| |:|:|
Db 132 LSPQIILQLG 142

RESULT 11
US-09-328-352-4571
; Sequence 4571, Application US/09328352
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4571
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4571

Query Match 62.7%; Score 37; DB 17; Length 291;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
| |:|:|
Db 169 NVTPEQVLEMEG 180

RESULT 12
US-09-417-507-30182
; Sequence 30182, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 30182
; LENGTH: 162
; TYPE: PRT
; ORGANISM: A.fumigatus
US-09-417-507-30182

Query Match 62.7%; Score 37; DB 23; Length 162;
Best Local Similarity 54.5%; Pred. No. 56;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 LKPEQILQLLG 12
| |:|:|
Db 42 LAPDSVLDLIG 52

RESULT 13
PCT-US94-00547-2
; Sequence 2, Application PC/TUS9400547
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00547
; FILING DATE: 18-JAN-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010409

;; FILING DATE: 29-JAN-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bastian, Kevin L.
;; REGISTRATION NUMBER: 34,774
;; REFERENCE/DOCKET NUMBER: 15280-113
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-543-9600
;; TELEFAX: 415-543-5043
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3135 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US94-00547-2

Query Match 61.0%; Score 36; DB 1; Length 3135;
Best Local Similarity 45.5%; Pred. No. 2.4e+03;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||||: |||:
Db 2284 NLKPKDVIEM 2294

RESULT 14
US-08-323-170-2
; Sequence 2, Application US/08323170
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Sulet 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/010,409
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 15280-113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-323-170-2

Query Match 61.0%; Score 36; DB 6; Length 3135;
Best Local Similarity 45.5%; Pred. No. 2.4e+03;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||||: |||:
Db 2284 NLKPKDVIEM 2294

RESULT 15
US-08-323-170A-2
; Sequence 2, Application US/08323170A
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170A
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 15280-113-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-323-170A-2

Query Match 61.0%; Score 36; DB 6; Length 3135;
Best Local Similarity 45.5%; Pred. No. 2.4e+03;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||||: |||:
Db 2284 NLKPKDVIEM 2294

Search completed: November 13, 1999, 05:08:42
Job time: 10868 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:08:08 : Search time 75.45 Seconds
(without alignments)
6.372 Million cell updates/sec

Title: US-08-913-430-14

Perfect score: 59

Sequence: 1 NLKPQIILQLG 12

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	66.1	399	D64327	H+-transporting ATP
2	37	62.7	484	A38905	dynein heavy chain
3	37	62.7	897	G03529	dynein heavy chain
4	37	62.7	2163	S50675	pre-mRNA splicing
5	37	62.7	55	PQ0827	E2/NS1 protein (CD
6	36	61.0	400	OKB02R	protein kinase (EC
7	36	61.0	262	PKBSK	kanamycin kinase (
8	36	61.0	1188	NDECKR	type I site-specif
9	36	61.0	623	JC8568	dihydrofolate redu
10	36	61.0	261	A26344	carbonate dehydrat
11	36	61.0	547	S70538	signal-transducin
12	36	61.0	283	A44814	endoflagellar prot
13	36	61.0	251	B64948	probable ABC trans
14	36	61.0	268	S75652	ABC-type transport
15	36	61.0	1170	A56157	chromosome segrega
16	36	61.0	3135	S37494	transmission block
17	36	61.0	658	S37494	squalene-hopene cy
18	36	61.0	2491	A57036	talin - slime mold
19	36	61.0	453	S59401	hypothetical prote
20	36	61.0	729	JC5812	norbin - rat
21	35	59.3	608	RDZOK1	dihydrofolate redu
22	35	59.3	418	OKHUR2	protein kinase (EC
23	35	59.3	415	OKRTR2	protein kinase (EC
24	35	59.3	418	OKBOR2	protein kinase (EC
25	35	59.3	608	G31262	dihydrofolate redu
26	35	59.3	610	A46005	dihydrofolate redu
27	35	59.3	1068	A43322	phosphatidylinosit
28	35	59.3	1068	I38110	1-phosphatidylinos
29	35	59.3	375	S75182	glutamate 5-kinase
30	35	59.3	1110	I59370	guanylate cyclase
31	35	59.3	306	F64185	D-alanine-D-alani
32	35	59.3	567	A42100	TGF-beta type II r
33	35	59.3	592	S51371	TGF-beta type II rece
34	35	59.3	567	JN0459	transforming growt
35	35	59.3	557	I50429	transforming growt
36	35	59.3	567	JC5372	transforming growt
37	35	59.3	478	JC5373	transforming growt
38	35	59.3	567	A44225	transforming growt
39	35	59.3	348	C64771	probable oxidoredu

UDP-MurNac-tripept
udp-n-acetylmuramy
hypothetical prote
hypothetical prote
SRG1 protein homol
serine/threonine-s

ALIGNMENTS

RESULT 1
D64327
H+-transporting ATP synthase (EC 3.6.1.34) subunit C - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Mar-1998
C:Accession: D64327
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999
A:Accession: D64327
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-399 <BUI>
A:Cross-references: GB:U67477; GB:L77117; NID:gl590949; PID:gl590956; TIGR:MJ0219; PI
C:Genetics:
A:Map position: REV210517-209318
C:Superfamily: H+ transporting ATP synthase C subunit
C:Keywords: hydrolase

Query Match 66.1%; Score 39; DB 2; Length 399;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NLKPQIILQLL 11
DB 386 NLKPPEIRSL 396
RESULT 2
A38905
dynein heavy chain, cytosolic - rat
N:Contains: dynein ATPase (EC 3.6.1.33)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Apr-1994 #sequence_revision 02-May-1994 #text_change 05-Sep-1997
C:Accession: A38905; I58139
R:Zhang, Z.; Tanaka, Y.; Nonaka, S.; Aizawa, H.; Kawasaki, H.; Nakata, T.; Hirokawa,
Proc. Natl. Acad. Sci. U.S.A. 90, 7928-7932, 1993
A:Title: The primary structure of rat brain (cytoplasmic) dynein heavy chain, a cytop
A:Reference number: A38905; MUID:93376715
A:Accession: A38905
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-4644 <ZHA>
A:Cross-references: GB:D13896; NID:g402527; PID:d1003502; PID:g402528
R:Mikami, A.; Paschal, B.M.; Mazumdar, M.; Vallee, R.B.
Neuron 10, 787-796, 1993
A:Title: Molecular cloning of the retrograde transport motor cytoplasmic dynein MAP 1
A:Reference number: I58139; MUID:93264075
A:Accession: I58139
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1023; 'MP', 1026-1771, 'D', 1773-2097, 'A', 2099-2138, 'V', 2140-2174, 'A', 2176-
44 <RES>
A:Cross-references: GB:L08505; NID:g294542; PID:g294543
C:Superfamily: dynein heavy chain, cytosolic
C:Keywords: ATP; blocked amino end; heterotetramer; hydrolase; microtubule binding; P
F:1904-1911/Region: nucleotide-binding motif A (P-loop)

F:2222-2229/Region: nucleotide-binding motif A (P-loop)
 F:2593-2600/Region: nucleotide-binding motif A (P-loop)
 F:2935-2942/Region: nucleotide-binding motif A (P-loop)
 F:1910/Binding site: ATP (lys) #status predicted
 F:2228/Binding site: ATP (lys) #status predicted
 F:2599/Binding site: ATP (lys) #status predicted
 F:2941/Binding site: ATP (lys) #status predicted

Query Match 62.7%; Score 37; DB 1; Length 4644;
 Best Local Similarity 63.8%; Pred. No. 4.3e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
 ||:|:|:|:|:|
 Db 1268 NLRPEALQAL 1278

RESULT 3

G02529
 dynein heavy chain 1, cytosolic - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
 C:Accession: G02529
 R:Vaisberg, E.A.; Grissom, P.M.; McIntosh, J.R.
 submitted to the EMBL Data Library, April 1996
 A:Reference number: H01399
 A:Accession: G02529
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-897 <VAI>
 A:Cross-references: EMBL:U53530; NID:gl314642; PID:gl314643
 C:Superfamily: dynein heavy chain, cytosolic

Query Match 62.7%; Score 37; DB 2; Length 897;
 Best Local Similarity 63.8%; Pred. No. 74;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
 ||:|:|:|:|:|
 Db 141 NLRPEALQAL 151

RESULT 4

S06675
 pre-mRNA splicing helicase BRR2 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: DNA repair protein RAD24; protein YER172c
 C:Species: Saccharomyces cerevisiae
 C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 05-Jun-1998
 C:Accession: S06675; S30856; JC4657
 R:Dietrich, F.S.
 submitted to the EMBL Data Library, December 1994
 A:Description: The sequence of S. cerevisiae cosmid 9163 and 9132.
 A:Reference number: S50428
 A:Accession: S06675
 A:Molecule type: DNA
 A:Residues: 1-2163 <DIE>
 A:Cross-references: EMBL:U18922; NID:g603405; PID:g603413; MIPS:YER172c
 R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor,
 submitted to the EMBL Data Library, February 1993
 A:Reference number: S30812
 A:Accession: S30856
 A:Molecule type: DNA
 A:Residues: 1-169 <MUL>
 A:Cross-references: EMBL:l11229
 R:Zhu, Y.B.; Prakash, L.; Prakash, S.
 Chinese Biochem. J. 11, 541-550, 1995
 A:Title: Molecular cloning and sequencing of DNA repair gene RAD24.
 A:Reference number: JC4657
 A:Accession: JC4657
 A:Molecule type: DNA
 A:Residues: 1-260, 'HCQT', 262, 'KT', 265 <ZHU>

C:Genetics:

A:Gene: SGD:BRR2; RAD24
 A:Cross-references: SGD:S0000974; MIPS:YER172c
 A:Map position: 5R
 C:Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology
 C:Keywords: ATP; DNA repair; nucleus; P-loop; pre-mRNA splicing
 F:521-877/Domain: DEAD/H box helicase homology <DEAD>
 F:521-528/Region: nucleotide-binding motif A (P-loop)
 F:630-635/Region: nucleotide-binding motif B
 F:634-637/Region: DEXH motif

Query Match 62.7%; Score 37; DB 2; Length 2163;
 Best Local Similarity 54.5%; Pred. No. 1.9e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 LKPEQILQLLG 12
 ||:|:|:|:|
 Db 863 LSPQDVLMGLG 873

RESULT 5

PQ0827
 E2/NS1 protein (cDNA) -- hepatitis C virus (strain NZL1) (fragments)
 C:Species: hepatitis C virus
 C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 14-Jul-1994
 C:Accession: PQ0827
 R:Okamoto, H.; Tokita, H.; Sakamoto, M.; Horikita, M.; Kojima, M.; Iizuka, H.; Mishir
 J. Gen. Virol. 74, 2385-2390, 1993
 A:Title: Characterization of the genomic sequence of type V (or 3a) hepatitis C virus
 A:Reference number: PQ0827
 A:Accession: PQ0827
 A:Molecule type: mRNA
 A:Residues: 1-55 <OKA>
 A:Cross-references: GB:D14305; GB:D14306
 C:Keywords: capsid protein; envelope protein; nonstructural protein

Query Match 62.7%; Score 37; DB 2; Length 55;
 Best Local Similarity 58.3%; Pred. No. 3.7;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
 ||:|:|:|:|
 Db 26 NMGPOQKLQLVG 37

RESULT 6

OR002R
 protein kinase (EC 2.7.1.37), CAMP-dependent, type II-alpha regulatory chain - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 15-Oct-1982 #sequence_revision 15-Oct-1982 #text_change 19-May-1995
 C:Accession: A00618; S17058
 R:Takio, K.; Smith, S.B.; Krebs, E.G.; Walsh, K.A.; Titani, K.
 Biochemistry 23, 4200-4206, 1984
 A:Title: Amino acid sequence of the regulatory subunit of bovine type II adenosine cy
 A:Reference number: A00618; MUID:85023306
 A:Accession: A00618
 A:Molecule type: protein
 A:Residues: 1-400 <TAK>
 R:Braun, R.K.; Vulliet, P.R.; Carbonaro-Hall, D.A.; Hall, F.L.
 Arch. Biochem. Biophys. 289, 187-191, 1991
 A:Title: Phosphorylation of RII subunit and attenuation of CAMP-dependent protein kin
 A:Reference number: S17058; MUID:91378531
 A:Accession: S17058
 A:Molecule type: protein
 A:Residues: 155-166 <BRA>
 C:Comment: The inactive form of the enzyme is composed of two regulatory chains and t
 our CAMP molecules
 C:Comment: Four types of regulatory chains are found: I-alpha, I-beta, II-alpha, and
 C:Comment: Type II regulatory chains are phosphorylated by the activated catalytic ch
 C:Comment: Type II regulatory chains mediate membrane association by binding to ancho
 merization.

C:Superfamily: CAMP-dependent protein kinase regulatory chain; cAMP receptor protein cyd
 C:Keywords: acetylated amino end; cAMP binding; duplication; heterotetramer; homodimer;
 F:1-134/Domain: protein interaction <DM>
 F:135-256/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <C1>
 F:257-389/Domain: cAMP receptor protein cyclic nucleotide-binding domain homology <C2>
 F:1/Modified site: acetylated amino end (Ser) #status experimental
 F:48,211/Binding site: phosphate (Thr) (covalent) (by proline-directed kinase) #status F
 F:95/Binding site: phosphate (Ser) (covalent) (by autophosphorylation) #status experimen
 F:204,213/Binding site: cAMP (Glu, Arg) #status predicted
 F:334,343/Binding site: cAMP (Glu, Arg) #status predicted

Query Match 61.0%; Score 36; DB 1; Length 400;
 Best Local Similarity 63.6%; Pred. No. 47;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
 |||||:|:|
 Db 138 NLDPEQLSQVL 148

RESULT 7

PKBSK
 Kanamycin kinase (EC 2.7.1.95) - Bacillus circulans
 N:Alternate names: aminoglycoside 3'-phosphotransferase; neomycin/kanamycin phosphotrans
 C:Species: Bacillus circulans
 C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 21-Nov-1998
 A:Accession: A00664; S10675
 R:Herbert, C.J.; Giles, I.G.; Akhtar, M.
 FEBS Lett. 160, 67-71, 1983
 A:Title: The sequence of an antibiotic resistance gene from an antibiotic-producing bact
 A:Reference number: A00664; MUID:83287774
 A:Accession: A00664
 A:Molecule type: DNA
 A:Residues: 1-262 <HER>
 A:Cross-references: EMBL:X03364
 R:Sawar, M.; Akhtar, M.
 Biochem. J. 268, 671-677, 1990
 A:Title: Cloning of aminoglycoside phosphotransferase (APH) gene from antibiotic- produc
 of the enzyme.
 A:Reference number: S10675; MUID:90303260
 A:Accession: S10675
 A:Molecule type: protein
 A:Residues: 'X',2,'QX',5-7,'X',9-12 <SAR>
 C:Superfamily: kanamycin kinase
 C:Keywords: antibiotic resistance; ATP; monomer; periplasmic space; phosphotransferase
 F:187/Active site: Asp #status predicted

Query Match 61.0%; Score 36; DB 1; Length 262;
 Best Local Similarity 66.7%; Pred. No. 30;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 PEQILQLL 12
 ||||:|:|
 Db 9 PEELELLG 17

RESULT 8

NDECKR
 Type I site-specific deoxyribonuclease (EC 3.1.21.3) EcoK chain R - Escherichia coli
 N:Alternate names: type I restriction enzyme EcoK R chain; type I restriction-modificati
 C:Species: Escherichia coli
 C:Date: 31-Mar-1990 #sequence_revision 26-Sep-1997 #text_change 15-May-1998
 A:Accession: S56576; H65249; S18776; A30375; Q00648
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
 Nucleic Acids Res. 23, 2105-2119, 1995
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
 A:Reference number: S56314; MUID:95334362
 A:Accession: S56576
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1188 <BUR>

A:Cross-references: EMBL:U14003; NID:g1263172; PID:g537192
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: H65249
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1188 <BLAT>
 A:Cross-references: GB:AE000505; GB:U00096; NID:g2367375; PID:g1790809; UWGP:b4350
 A:Experimental source: strain K-12, substrain MG1655
 R:Waite-Rees, P.A.; Keating, C.J.; Moran, L.S.; Slatko, B.E.; Hornstra, L.J.; Benner,
 J. Bacteriol. 173, 5207-5219, 1991
 A:Title: Characterization and expression of the Escherichia coli Mrr restriction syst
 A:Reference number: A40368; MUID:91317743
 A:Accession: S18776
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-27 <WAI>
 A:Cross-references: EMBL:X54198
 R:Loenen, W.A.M.; Daniel, A.S.; Braymer, H.D.; Murray, N.E.
 J. Mol. Biol. 198, 159-170, 1987
 A:Title: Organization and sequence of the hsd genes of Escherichia coli K-12.
 A:Reference number: A30375; MUID:88118919
 A:Accession: A30375
 A:Molecule type: DNA
 A:Residues: 1-628, ECGKNSHRHPGATCAGPRAGLPLYLPRYRFRSDRPGSAYSDHHPQAGGUSLQRRAGRAH
 A:Cross-references: GB:X06545; NID:g41751; PID:g41752
 C:Comment: This is one of three components (S, R, and M chains) of type I site-specif
 ctors: it is the site-specificity determinant for the catalytic activities of the enz
 C:Genetics:
 A:Gene: hsdR
 A:Map position: 99 min
 C:Superfamily: type I site-specific deoxyribonuclease EcoK chain R; DEAD/H box helica
 C:Keywords: ATP; DNA binding; hydrolase; P-loop; restriction modification system
 F:489-847/Domain: DEAD/H box helix homology <DEAD>
 F:489-496/Region: nucleotide-binding motif A (P-loop)
 F:588-593/Region: nucleotide-binding motif B
 F:592-595/Region: DEAH motif

Query Match 61.0%; Score 36; DB 1; Length 1188;
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 3 RPEQILQLL 12
 ||||:|:|
 Db 433 RPELEMLG 442

RESULT 9

JC6568
 dihydrofolate reductase (EC 1.5.1.3) / thymidylate synthase (EC 2.1.1.45) - Plasmodiu
 C:Species: Plasmodium vivax
 C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 01-Feb-1999
 A:Accession: JC6568
 R:de Pecoulas, P.E.; Basco, L.K.; Tahar, R.; Ouatas, T.; Mazabraud, A.
 Gene 211, 177-185, 1998
 A:Title: Analysis of the Plasmodium vivax dihydrofolate reductase-thymidylate synthas
 A:Reference number: JC6568
 A:Accession: JC6568
 A:Molecule type: DNA
 A:Residues: 1-623 <DEP>
 A:Cross-references: EMBL:X98123
 C:Genetics:
 A:Gene: dhfr-ts
 C:Superfamily: bifunctional dihydrofolate reductase--thymidylate synthase; thymidylat
 C:Keywords: NADP; oxidoreductase
 F:340-623/Domain: thymidylate synthase homology <TDS>

Query Match 61.0%; Score 36; DB 2; Length 623;
Best Local Similarity 50.0%; Pred. No. 76;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
||:|:|:|
Db 552 NLQPAQFIHLG 563

RESULT 10
A26344
carbonate dehydratase (EC 4.2.1.1) I - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 25-Apr-1997
C:Accession: A26344; I49573
R:Fraser, P.J.; Curtis, P.J.
J. Mol. Evol. 23, 294-299, 1986
A:Title: Molecular evolution of the carbonic anhydrase genes: calculation of divergence
A:Reference number: A26344; MUID:87169766
A:Accession: A26344
A:Molecule type: mRNA
A:Residues: 1-261 <FRA>
R:Fraser, P.; Cummings, P.; Curtis, P.
Mol. Cell. Biol. 9, 3308-3313, 1989
A:Title: The mouse carbonic anhydrase I gene contains two tissue-specific promoters.
A:Reference number: I49573; MUID:90014784
A:Accession: I49573
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-82, 'G', 83-100, 102-261 <RES>
A:Cross-references: GB:L36655; NID:9556293; PID:9556295
C:Genetics:
A:Gene: Car1
A:Introns: 13/1; 79/1; 118/3; 150/3; 171/3; 223/3
C:Superfamily: carbonate dehydratase; carbonic anhydrase homology
C:Keywords: carbon-oxygen lyase; hydro-lyase; zinc
F: 6-261/Domain: carbonic anhydrase homology <CAH>

Query Match 61.0%; Score 36; DB 2; Length 261;
Best Local Similarity 58.3%; Pred. No. 30;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
||:|:|:|
Db 218 SLSPEQLAQLRG 229

RESULT 11
S70538
signal-transducing histidine kinase cita - Klebsiella pneumoniae
N:Alternate names: sensor kinase
C:Species: Klebsiella pneumoniae
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Aug-1998
C:Accession: S70538; S70537
R: Bott, M.
submitted to the EMBL Data Library, July 1995
A:Reference number: S70538
A:Accession: S70538
A:Molecule type: DNA
A:Residues: 1-547 <BOT>
A:Cross-references: EMBL:U31454; NID:9924990; PID:9924991
A:Experimental source: ATCC 13882
R: Bott, M.; Meyer, M.; Dimroth, P.
Mol. Microbiol. 18, 533-546, 1995
A:Title: Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae.
A:Reference number: S70535; MUID:96342382
A:Accession: S70537
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 20-48; 173-176; 178-199; 202-205; 214-229; 316-331; 348-356; 436-447; 476-483; 491-495
A:Cross-references: EMBL:U31464
A:Experimental source: ATCC 13882

C:Genetics:
A:Gene: cita
C:Superfamily: two-component sensor histidine kinase; sensor histidine kinase homolog
F:24-44/Domain: transmembrane #status predicted <TML>
F:45-177/Domain: periplasmic #status predicted <PER>
F:178-199/Domain: transmembrane #status predicted <TM2>
F:350/Binding site: phosphate (His) (covalent) (by autophosphorylation) #status predicted

Query Match 61.0%; Score 36; DB 2; Length 547;
Best Local Similarity 54.5%; Pred. No. 66;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 11
||:|:|:|
Db 214 NMEPQOLSOLL 224

RESULT 12
A44814
endoflagellar protein ftaB - Leptospira borgpetersenii
C:Species: Leptospira borgpetersenii
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 03-Feb-1994
C:Accession: A44814
R:Michelson, M.; Rood, J.I.; Faine, S.; Adler, B.
J. Gen. Microbiol. 137, 1529-1536, 1991
A:Title: Molecular analysis of a Leptospira borgpetersenii gene encoding an endoflagellar protein.
A:Reference number: A44814; MUID:92065222
A:Accession: A44814
A:Molecule type: DNA
A:Residues: 1-283 <MIT>
A:Note: sequence extracted from NCBI backbone (NCBIN:66827, NCBIP:66829)
C:Genetics:
A:Gene: ftaB
C:Superfamily: flagellin

Query Match 61.0%; Score 36; DB 2; Length 283;
Best Local Similarity 54.5%; Pred. No. 33;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 11
||:|:|:|
Db 270 NKKPNSVLKLL 280

RESULT 13
B64948
probable ABC transport system ATP-binding protein yebM - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 05-Dec-1998
C:Accession: B64948
R:Blattner, F.R.; Mau, B.; Shao, Y.
A:Rose, D.J.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: B64948
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-251 <BLAT>
A:Cross-references: GB:AE000280; GB:U00096; NID:q1788163; PID:q1788165; UWGP:bl858
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yebM
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
C:Keywords: ATP; P-loop; transport protein
F:20-197/Domain: ATP-binding cassette homology <ABC>
F:37-44/Region: nucleotide-binding motif A (P-loop)

Query Match 61.0%; Score 36; DB 2; Length 251;

Best Local Similarity 72.7%; Pred. No. 29;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 LKPEQILQLLG 12
||| :|||
Db 27 LKPGKILTLG 37

RESULT 14
S75652
ABC-type transport protein sl11878 - *Synechocystis* sp. (strain PCC 6803)
N:Alternate names: protein sl11878
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S75652
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-368 <KAN>
A:Cross-references: EMBL:D90912; GB:AB001339; NID:g1653228; PID:d1018946; PID:g1653298
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: P-loop
F:36-233/Domain: ATP-binding cassette homology <ABC>
F:53-60/Region: nucleotide-binding motif A (P-loop)

Query Match 61.0%; Score 36; DB 2; Length 368;
Best Local Similarity 66.7%; Pred. NO. 43;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
||| :|||
Db 42 NLAPGEILGLG 53

RESULT 15
A56157
Chromosome segregation protein SMC2 - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: DA-box protein; protein R003; protein YFR031c; structural maintenance
C:Species: *Saccharomyces cerevisiae*
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 17-Mar-1999
C:Accession: A56157; S56286; S62242; S63836; S48530
R:Strunnikov, A.V.; Hogan, E.; Koshland, D.
Genes Dev. 9, 587-599, 1995
A:Title: SMC2, a *Saccharomyces cerevisiae* gene essential for chromosome segregation and
A:Reference number: A56157; MUID:95212908
A:Accession: A56157
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1170 <STR>
A:Cross-references: GB:U05820; NID:g468039; PID:g468040
R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanu
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces* ce
A:Reference number: S56186
A:Accession: S56286
A:Molecule type: DNA
A:Residues: 1-1170 <MUR>
A:Cross-references: EMBL:D50617; NID:g836685; PID:d1009911; PID:g836786; MIPS:YFR031C
R:Murakami, Y.
submitted to the EMBL Data Library, December 1994
A:Reference number: S62230
A:Accession: S62242
A:Molecule type: DNA
A:Residues: 1-1170 <MUW>

A:Cross-references: EMBL:D44602; NID:g893419; PID:d1008633; PID:g893426
R:Eki, T.; Naitou, M.; Hagiwara, H.; Abe, M.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, M.
Yeast 12, 177-190, 1996
A:Title: Fifteen open reading frames in a 30.8 kb region of the right arm of chromosome
A:Reference number: S63830; MUID:96287654
A:Accession: S63836
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1170 <EKI>
A:Cross-references: EMBL:D44602; NID:g893419; PID:d1008633; PID:g893426
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1995
C:Genetics:
A:Gene: SGD:SMC2
A:Cross-references: SGD:S0001927; MIPS:YFR031C
A:Map position: 6R
C:Superfamily: conserved hypothetical P115 protein
C:Keywords: transmembrane protein
F:41-57/Domain: transmembrane #status predicted <TM1>
F:1090-1106/Domain: transmembrane #status predicted <TM2>

Query Match 61.0%; Score 36; DB 2; Length 1170;
Best Local Similarity 54.5%; Pred. NO. 1.5e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||| :|||
Db 155 NKKPSEILSLI 165

Search completed: November 13, 1999, 12:08:09
Job time: 2081 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	39	66.1	399	1	ATPC_METJA	Q57672	methanococ
2	37	62.7	2163	1	BRR2_YEAST	P32639	saccharomyc
3	37	62.7	4644	1	DYHC_RAT	P38650	rattus norv
4	36	61.0	260	1	CAHL_MOUSE	P13634	mus musculu
5	36	61.0	547	1	CITA_KLEPN	P52687	klebsiella
6	36	61.0	616	1	DRTS_PLAVI	Q02604	plasmodium
7	36	61.0	400	1	KAP2_BOVIN	P00515	bos taurus
8	36	61.0	262	1	KK4A_BACCI	P00553	bacillus ci
9	36	61.0	3135	1	S220_PLAFO	Q08372	plasmodium
10	36	61.0	1170	1	SMCZ_YEAST	P38989	saccharomyc
11	36	61.0	658	1	SHQC_ZYMMO	P33990	zymomonas m
12	36	61.0	1188	1	T1R_ECOLI	P08956	escherichia
13	36	61.0	2491	1	TALA_DICDI	P54633	dictyosteli
14	36	61.0	251	1	ZNVC_ECOLI	P52648	escherichia
15	35	59.3	477	1	ARP7_YEAST	Q12406	saccharomyc
16	35	59.3	1110	1	CYCX_RAT	P51839	rattus norv
17	35	59.3	306	1	DDL_HAEIN	P44405	haemophilus
18	35	59.3	608	1	DRTS_PLAFK	P13922	plasmodium
19	35	59.3	610	1	DRTS_TOXGO	Q07422	toxoplasma
20	35	59.3	417	1	KAP3_BOVIN	P31322	bos taurus
21	35	59.3	417	1	KAP3_HUMAN	P31323	homo sapien
22	35	59.3	415	1	KAP3_RAT	P12369	rattus norv
23	35	59.3	689	1	KELC_DROME	Q04652	drosophilla
24	35	59.3	486	1	NAM9_YEAST	P27929	saccharomyc
25	35	59.3	1068	1	P11A_BOVIN	P32871	bos taurus
26	35	59.3	1068	1	P11A_HUMAN	P42336	homo sapien
27	35	59.3	1068	1	P11A_MOUSE	P42337	mus musculu
28	35	59.3	375	1	PROB_SYNY3	P73071	synecocyst
29	35	59.3	567	1	TGR2_HUMAN	P37173	homo sapien
30	35	59.3	592	1	TGR2_MOUSE	Q62312	mus musculu
31	35	59.3	567	1	TGR2_RAT	P38438	rattus norv
32	35	59.3	267	1	Y125_MYCCA	P53661	mycoplasma
33	35	59.3	324	1	YAJ0_ECOLI	P77735	escherichia
34	35	59.3	942	1	ZDS2_YEAST	P54786	saccharomyc
35	34.5	58.5	3415	1	POIG_POWL	Q04538	t genome po
36	34	57.6	492	1	ADRO_BOVIN	P08165	bos taurus
37	34	57.6	794	1	CADC_HUMAN	P55289	homo sapien
38	34	57.6	423	1	CYAA_STIAU	P40137	stigmatella
39	34	57.6	583	1	ESTR_OREAU	P50240	oreochromis
40	34	57.6	304	1	FLA_BACSU	P02968	calliculus su
41	34	57.6	602	1	H0XF_ALCEU	P22317	acalligenes
42	34	57.6	817	1	HUNB_MUSDO	Q01778	musca domes
43	34	57.6	1025	1	IT48_HUMAN	P53708	homo sapien

OC SACCHAROMYCES CEREVISIAE (BARKER & LEAS).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.

RA SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA DITRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNNICKE-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.,
RL SUBMITTED (DEC-1994) TO ENBL/GENBANK/DBBJ DATA BANKS.
[2]
RN SEQUENCE OF 1-169 FROM N.A.
RA MULLIGAN J.T., DITRICH F.S., HENNESSEY K.M., SEHL P., KOMP C.,
RA WEI Y., TAYLOR P., NAKAHARA K., ROBERTS D., DAVIS R.W.,
RL SUBMITTED (FEB-1993) TO ENBL/GENBANK/DBBJ DATA BANKS.
[3]
RN CHARACTERIZATION.
RC STRAIN=DRY473;
RX MEDLINE; 96304576.
RA NOBLE S.M., GUTHRIE C.;
RT "Identification of novel genes required for yeast pre-mRNA splicing
by means of cold-sensitive mutations.";
RL GENETICS 143:67-80(1996).
CC -1- FUNCTION: REQUIRED FOR PRE-MRNA SPLICING.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SKI2 SUBFAMILY OF HELICASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; U18922; G603413; -
DR PIR; S30856; S30856.
DR SGD; L0003100; BRR2.
DR PFAM; PF00270; DEAD; 1.
DR PFAM; PF00271; helicase C; 1.
KW HELICASE; ATP-BINDING; NUCLEAR PROTEIN; MRNA PROCESSING.
FT NP_BIND 79 86 ATP (POTENTIAL).
FT NP_BIND 521 528 ATP (POTENTIAL).
FT SITE 634 637 DEH BOX.
FT SITE 2163 637 DEH BOX.
SQ SEQUENCE 2163 AA; 246183 MW; 48DB127B CRC32;

Query Match 62.7%; Score 37; DB 1; Length 2163;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 LKPEQILQLLG 12
| | | | | | | | | |
Db 863 LSPQDVLQMLG 873

RESULT 3
DYHC_RAT STANDARD; PRT; 4644 AA.
AC P38650; Q63178;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC) (MAP 1C).
GN DNCH1 OR DNCHL OR MAPIC.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=BRAIN;
RX MEDLINE; 93376715.

RA ZHANG Z., TANAKA Y., NONAKA S., AIZAWA H., KAWASAKI H., NAKATA T.,
RA HIROKAWA N.;
RT "The primary structure of rat brain (cytoplasmic) dynein heavy chain,
a cytoplasmic motor enzyme.";
RL PROC. NATL. ACAD. SCI. U.S.A. 90:7928-7932(1993).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE; 93264075.
RA MIKAMI A., PASCHAL B.M., MAZUMDAR M., VALLEE R.B.;
RT "Molecular cloning of the retrograde transport motor cytoplasmic
dynein (MAP 1C).";
RL NEURON 10:787-796(1993).
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
ORGANELLES ALONG MICROTUBULES.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
INTERMEDIATE AND LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; D13896; G402528; -
DR EMBL; L08505; G294543; -
DR PIR; A38905; A38905.
KW MOTOR PROTEIN; MICROTUBULES; DYNEIN; ATP-BINDING;
FT NP_BIND 1904 1911 ATP (POTENTIAL).
FT NP_BIND 2222 2229 ATP (POTENTIAL).
FT NP_BIND 2593 2600 ATP (POTENTIAL).
FT NP_BIND 2935 2942 ATP (POTENTIAL).
FT CONFLICT 1024 1025 SR -> MP (IN REF. 2).
FT CONFLICT 1772 1772 N -> D (IN REF. 2).
FT CONFLICT 2098 2098 P -> A (IN REF. 2).
FT CONFLICT 2139 2139 F -> V (IN REF. 2).
FT CONFLICT 2175 2175 D -> A (IN REF. 2).
FT CONFLICT 2185 2185 K -> Q (IN REF. 2).
FT CONFLICT 2366 2366 L -> V (IN REF. 2).
FT CONFLICT 2382 2382 T -> S (IN REF. 2).
FT CONFLICT 2463 2463 G -> A (IN REF. 2).
FT CONFLICT 3219 3219 A -> D (IN REF. 2).
FT CONFLICT 4131 4131 R -> K (IN REF. 2).
FT CONFLICT 4366 4366 F -> S (IN REF. 2).
FT CONFLICT 4511 4511 A -> G (IN REF. 2).
SQ SEQUENCE 4644 AA; 532240 MW; A3E60EF3 CRC32;

Query Match 62.7%; Score 37; DB 1; Length 4644;
Best Local Similarity 63.6%; Pred. No. 2.7e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
| | | | | | | | | |
Db 1268 NLRPEALQAL 1278

RESULT 4
CAHL_MOUSE STANDARD; PRT; 260 AA.
ID CAHL_MOUSE
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE CARBONIC ANHYDRASE I (EC 4.2.1.1) (CARBONATE DEHYDRATASE I).
GN CAI OR CARL.
OS MUS MUSCULUS (MOUSE).

```

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC KLEBSIELLA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 13882;
RX MEDLINE; 96342382.
RT BOTT M., MEYER M., DIMROTH P.;
RT "Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae.";
RL MOL. MICROBIOL. 18:533-546(1995).
CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM CITA/CITB
CC ESSENTIAL FOR EXPRESSION OF CITRATE-SPECIFIC FERMENTATION GENES.
CC MAY ACTIVATE CITB BY PHOSPHORYLATION.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U31464; G924991; -.
CC PFAM; PF00512; signal; 1.
CC SENSORY TRANSDUCTION; TRANSFERASE; KINASE; PHOSPHORYLATION;
CC TRANSMEMBRANE; INNER MEMBRANE.
CC FT DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 24 44 POTENTIAL.
CC FT DOMAIN 45 180 PERIPLASMIC (POTENTIAL).
CC FT TRANSMEM 181 201 POTENTIAL.
CC FT DOMAIN 202 547 CYTOPLASMIC (POTENTIAL).
CC MOD_RES 350 350 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC SQ SEQUENCE 547 AA; 61780 MW; BC0FD2BB CRC32;
Query Match 61.0%; Score 36; DB 1; Length 547;
Best Local Similarity 54.5%; Pred.No. 38;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 NLRPEQILQLL 11
Db 214 NMSPQQLSLL 224
:::1:1:111
RESULT 6
DRTS_PLAVI STANDARD; PRT; 616 AA.
AC 002604;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DIHYDROFOLATE REDUCTASE (EC 1.5.1.3) / THYMIDYLATE SYNTHASE
DE (EC 2.1.1.45) (DHFR-TS) (FRAGMENT).
DE PLASMODIUM VIVAX.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RN SEQUENCE FROM N.A.
RL ELDIN DE PECOULAS P.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: 5,6,7,8-TETRAHYDROFOLATE + NADP(+) -
CC 7,8-DIHYDROFOLATE + NADPH.
CC -!- CATALYTIC ACTIVITY: 5,10-METHYLENETETRAHYDROFOLATE + DUMP -
CC DIHYDROFOLATE + DTMP.
CC -!- PATHWAY: ESSENTIAL STEP FOR DE NOVO GLYCINE AND PURINE SYNTHESIS,
CC DNA PRECURSOR SYNTHESIS, AND FOR THE CONVERSION OF DUMP TO DTMP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X98123; E248438; -;
DR PROSITE; PS00075; DHFR; 1.
DR PROSITE; PS00091; THYMIDYLATE SYNTHASE; 1.
DR PFAM; PF00186; Dihfolate_red; 1.
DR PFAM; PF00303; thymidylat_synt; 1.
DR HSP; P00470; IAN5
DR MULTIFUNCTIONAL ENZYME; OXIDOREDUCTASE; TRANSFERASE; NADP;
KW METHYLTRANSFERASE; NUCLEOTIDE BIOSYNTHESIS; ONE-CARBON METABOLISM.
FT NON_TER 1 1
FT DOMAIN <1 233 DIHYDROFOLATE REDUCTASE.
FT DOMAIN 330 616 THYMIDYLATE SYNTHASE.
FT ACT_SITE 498 498 BY SIMILARITY..
SQ SEQUENCE 616 AA; 70252 MW; 5F5DCC16 CRC32;

Query Match 61.0%; Score 36; DB 1; Length 616;
Best Local Similarity 50.08; Pred. No. 44;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
||| | : ||
Db 545 NLQPAQFIHLG 556

RESULT 7
ID KAP2_BOVIN STANDARD; PRT; 400 AA.
AC P00515;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE CAMP-DEPENDENT PROTEIN KINASE TYPE II-ALPHA REGULATORY CHAIN.
GN PKR2A.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE.
RC TISSUE=HEART MUSCLE;
RX MEDLINE; 82222175.
RA TAKIO K., SMITH S.B., KREBS E.G., WALSH K.A., TITANI K.;
RT "Primary structure of the regulatory subunit of type II
RT CAMP-dependent protein kinase from bovine cardiac muscle."
RL PROC. NATL. ACAD. SCI. U.S.A. 79:2544-2548(1982).
RN [2]
RP 3D-STRUCTURE MODELLING.
RX MEDLINE; 87157645.
RA WEBER I.T., STEITZ T.A., BUBIS J., TAYLOR S.S.;
RT "Predicted structures of CAMP binding domains of type I and II
RT regulatory subunits of CAMP-dependent protein kinase."
RL BIOCHEMISTRY 26:343-351(1987).
CC -1- SUBUNIT: THE INACTIVE FORM OF THE ENZYME IS COMPOSED OF TWO
CC REGULATORY CHAINS AND TWO CATALYTIC CHAINS. ACTIVATION BY CAMP
CC PRODUCES TWO ACTIVE CATALYTIC MONOMERS AND A REGULATORY DIMER
CC THAT BINDS FOUR CAMP MOLECULES.
CC -1- TISSUE SPECIFICITY: FOUR TYPES OF REGULATORY CHAINS ARE FOUND:
CC I-ALPHA, I-BETA, II-ALPHA, AND II-BETA. THEIR EXPRESSION VARIES
CC AMONG TISSUES AND IS IN SOME CASES CONSTITUTIVE AND IN OTHERS
CC INDUCIBLE.
CC -1- PTM: A SECOND PHOSPHORYLATION SITE HAS NOT BEEN LOCATED.
DR PIR; A0618; OR02R.
DR PIR; S17058; S17058.
DR PDB; 2APK; 15-OCT-94.
DR PDB; 2BPK; 15-OCT-94.
DR PROSITE; PS00888; CNMP_BINDING_1; 2.
DR PROSITE; PS00889; CNMP_BINDING_2; 2.
DR PROSITE; PS50042; CNMP_BINDING_3; 2.
DR PFAM; PF00027; CNMP_binding; 2.

KW CAMP-BINDING; PHOSPHORYLATION; DUPLICATION; MULTIGENE FAMILY;
KW ACETYLATION; 3D-STRUCTURE.
FT DOMAIN 1 134 DIMERIZATION AND PHOSPHORYLATION.
FT NP_BIND 135 256 CAMP.
FT NP_BIND 257 400 CAMP.
FT REPEAT 135 256
FT REPEAT 257 400
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 95 95 PHOSPHORYLATION (BY ACTIVATED ENZYME).
FT BINDING 204 204 CAMP.
FT BINDING 213 213 CAMP.
FT BINDING 334 334 CAMP.
FT BINDING 343 343 CAMP.
SQ SEQUENCE 400 AA; 44962 MW; 1F7A0D99 CRC32;

Query Match 61.0%; Score 36; DB 1; Length 400;
Best Local Similarity 63.6%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
||| | : ||
Db 138 NLDPEQLSQVL 148

RESULT 8
ID KKA4_BACCI STANDARD; PRT; 262 AA.
AC P00553;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE AMINOGLYCOSIDE 3'-PHOSPHOTRANSFERASE (EC 2.7.1.95) (KANAMYCIN KINASE,
DE TYPE IV) (NEOMYCIN-KANAMYCIN PHOSPHOTRANSFERASE, TYPE IV) (APH(3')IV)
DE (BUTIROIN RESISTANCE PROTEIN).
GN BUTA OR APH OR APHA4.
OS BACILLUS CIRCULANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86158748.
RA HERBERT C.J., SARWAR M., NER S.S., GILES I.G., AKHTAR M.;
RT "Sequence and interspecies transfer of an aminoglycoside
RT phosphotransferase gene (APH) of Bacillus circulans. Self-defence
RT mechanism in antibiotic-producing organisms."
RL BIOCHEM. J. 233:383-393(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 83287774.
RA HERBERT C.J., GILES I.G., AKHTAR M.;
RT "The sequence of an antibiotic resistance gene from an antibiotic-
RT producing bacterium. Homologies with transposon genes."
RL FEBS LETT. 160:67-71(1983).
RN [3]
RP SEQUENCE OF 1-25 FROM N.A.
RC STRAIN-NRRL B3312;
RX MEDLINE; 94374689.
RA AUBERT-PIVET E., DAVIES J.;
RT "Biosynthesis of butirosin in Bacillus circulans NRRL B3312:
RT identification by sequence analysis and insertional mutagenesis of
RT the butB gene involved in antibiotic production."
RL GENE 147:1-11(1994).
RN [4]
RP PARTIAL SEQUENCE OF 1-12.
RX MEDLINE; 90303260.
RA SARWAR M., AKHTAR M.;
RT "Cloning of aminoglycoside phosphotransferase (APH) gene from
RT antibiotic-producing strain of Bacillus circulans into a
RT high-expression vector, pKK223-3. Purification, properties and
RT location of the enzyme."
RL BIOCHEM. J. 268:671-677(1990).
CC -1- FUNCTION: RESISTANCE TO BUTIROIN AND STRUCTURALLY-RELATED

CC AMINOGLYCOSIDES, INCLUDING KANAMYCIN AND AMIKACIN.
 CC -!- CATALYTIC ACTIVITY: ATP + KANAMYCIN = ADP + KANAMYCIN
 CC 3'-PHOSPHATE (ALSO ACTS ON OTHER ANTIBIOTICS).
 CC -!- SUBUNIT: MONOMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: TO OTHER AMINOGLYCOSIDE PHOSPHOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X03364; G39414; -.
 CC EMBL; L20421; G688437; -.
 CC PIR; A00664; PKBSK.
 KW ANTIBIOTIC RESISTANCE; TRANSFERASE; KINASE; ATP-BINDING.
 FT ACT-SITE 187 BY SIMILARITY.
 FT CONFLICT 3 3 E -> Q (IN REF. 4).
 FT CONFLICT 106 106 K -> N (IN REF. 2).
 SQ SEQUENCE 262 AA; 29913 MW; 94DBA99E CRC32;

 Query Match 61.0%; Score 36; DB 1; Length 262;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 4 PEQIQLLG 12
 DB 9 PEELLELLG 17
 |||:|:|
 |||:|:|

 RESULT 9
 S230_PLAFO STANDARD; PRT; 3135 AA.
 ID S230_PLAFO STANDARD; PRT; 3135 AA.
 AC Q08372;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE TRANSMISSION-BLOCKING TARGET ANTIGEN S230 PRECURSOR.
 GN S230.
 OS PLASMODIUM FALCIPARUM (ISOLATE NF54), AND
 OS PLASMODIUM FALCIPARUM (ISOLATE 3D7).
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ISOLATE NF54;
 RX MEDLINE; 93241227.
 RA WILLIAMSON K.C., CRISCO M.D., KASLOW D.C.;
 RT "Cloning and expression of the gene for Plasmodium falciparum
 RT transmission-blocking target antigen, Pf230.";
 RL MOL. BIOCHEM. PARASITOL. 58:355-358(1993).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE 3D7;
 RA BHATTI S., ALANO P., LUO C., HANSRA S., AIKAWA M., CARTER R.,
 RA ELLIOTT J.F.;
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- SUBCELLULAR LOCATION: SURFACE PROTEIN.
 CC -!- DEVELOPMENTAL STAGE: GAMETOCYTE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L08135; G294176; -.
 CC EMBL; L04162; G605631; -.

DR PIR; A48584; A48584.
 KW MEMBRANE; REPEAT; MALARIA; ANTIGEN; SIGNAL.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 3135 TRANSMISSION BLOCKING TARGET ANTIGEN
 FT S230.
 FT DOMAIN 280 304 POLY-GLU.
 FT DOMAIN 379 410 8 X 4 AA TANDEM REPEATS OF E-E-V-G.
 FT REPEAT 379 382 1-1.
 FT REPEAT 383 386 1-2.
 FT REPEAT 387 390 1-3.
 FT REPEAT 391 394 1-4.
 FT REPEAT 395 398 1-5.
 FT REPEAT 399 402 1-6.
 FT REPEAT 403 406 1-7.
 FT REPEAT 407 410 1-8.
 FT DOMAIN 411 442 4 X 8 AA APPROXIMATE TANDEM REPEATS OF
 FT E-E-V-G-E-[EG]-[EV]-G.
 FT REPEAT 411 418 2-1.
 FT REPEAT 419 426 2-2.
 FT REPEAT 427 434 2-3.
 FT REPEAT 435 442 2-4.
 SQ SEQUENCE 3135 AA; 363213 MW; FCAEF748 CRC32;

 Query Match 61.0%; Score 36; DB 1; Length 3135;
 Best Local Similarity 45.5%; Pred. No. 2.6e+02;
 Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NLKPEQLQLL 11
 DB 2284 NLKPKDVIELM 2294
 ||||:|:|:
 ||||:|:|:

 RESULT 10
 SMC2_YEAST STANDARD; PRT; 1170 AA.
 ID SMC2_YEAST STANDARD; PRT; 1170 AA.
 AC P38989;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE CHROMOSOME SEGREGATION PROTEIN SMC2 (DA-BOX PROTEIN SMC2).
 GN SMC2 OR YFR031C.
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
 OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95212908.
 RA STRUNNIKOV A.V., HOGAN E., KOSHLAND D.;
 RT "SMC2, a Saccharomyces cerevisiae gene essential for chromosome
 RT segregation and condensation, defines a subgroup within the SMC
 RT family.";
 RL GENES DEV. 9:587-599(1995).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE; 95400292.
 RA MURAKAMI Y., NAITOU M., HAGIWARA H., SHIBATA T., OZAWA M.,
 RA SASANUMA S.-I., SASANUMA M., TSUCHIYA Y., SOEDA E., YOKOYAMA K.,
 RA YAMAZAKI M., TASHIRO H., EKI T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RT Saccharomyces cerevisiae.";
 RL NAT. GENET. 10:261-268(1995).
 RL [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE; 96287654.
 RA EKI T., NAITOU M., HAGIWARA H., ABE M., OZAWA M., SASANUMA S.-I.,
 RA SASANUMA M., TSUCHIYA Y., SHIBATA T., WTANABE K., ONO A.,
 RA YAMAZAKI M.-A., TASHIRO H., HANAOKA F., MURAKAMI Y.;
 RT "Fifteen open reading frames in a 30.8 kb region of the right arm of
 RT chromosome VI from Saccharomyces cerevisiae.";
 RL YEAST 12:177-190(1996).
 RL

CC -!- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC1 OR OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U05820; G468040; -
CC EMBL: D50617; D1009911; -
CC EMBL: D44602; D1008633; -
CC PIR: S48530; S48530.
CC SGD: L0001927; SMC2.
CC HSP: P02633; IBD.
CC MITOSIS; ATP-BINDING; COILED COIL; NUCLEAR PROTEIN.
CC NP_BIND 32 39
CC FT DOMAIN 172 469
CC FT DOMAIN 678 1027
CC FT DOMAIN 1093 1119
CC FT ALA/ASP-RICH (DA-BOX).
CC SQ SEQUENCE 1170 AA; 133927 MW; C8865AC0 CRC32;

Query Match 61.0%; Score 36; DB 1; Length 1170;
Best Local Similarity 54.5%; Pred. No. 89;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
I::I::I::I::
Db 155 NMPKSEILSLI 165

RESULT 11
SOHC_ZYMO
ID SOHC-ZYMO STANDARD; PRT; 658 AA.
AC P33990;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SQUALENE--HOPENE CYCLASE (EC 5.4.99.-).
GN SHC.
OS ZYMONAS MOBILIS.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; ZYMONAS GROUP;
OC ZYMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 31821 / CP4;
RX MEDLINE; 95202074.
RA REIPEN I.G., PORALIA K., SAHM H., SPRENGER G.A.;
RT "Zymonas mobilis squalene-hopene cyclase gene (shc): cloning, DNA
RL sequence analysis, and expression in Escherichia coli.";
RL MICROBIOLOGY 141:155-161(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 31821 / CP4;
RX MEDLINE; 95202074.
RA REIPEN I.G., PORALIA K., SAHM H., SPRENGER G.A.;
RT "Zymonas mobilis squalene-hopene cyclase gene (shc): cloning, DNA
RL sequence analysis, and expression in Escherichia coli.";
RL MICROBIOLOGY 141:155-161(1995).
RN [2]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U05820; G468040; -
CC EMBL: D50617; D1009911; -
CC EMBL: D44602; D1008633; -
CC PIR: S48530; S48530.
CC SGD: L0001927; SMC2.
CC HSP: P02633; IBD.
CC MITOSIS; ATP-BINDING; COILED COIL; NUCLEAR PROTEIN.
CC NP_BIND 32 39
CC FT DOMAIN 172 469
CC FT DOMAIN 678 1027
CC FT DOMAIN 1093 1119
CC FT ALA/ASP-RICH (DA-BOX).
CC SQ SEQUENCE 1170 AA; 133927 MW; C8865AC0 CRC32;

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X80766; G677871; -
CC EMBL: X73561; G405608; -
CC EMBL: AJ001401; E1169922; -
CC PIR: S37494; S37494.
CC PROSITE: PS01074; TERPENE SYNTHASES; 1.
CC PFAM: PF00432; prenyltrans; 3.
CC HSP: P33247; LSQC.
CC ISOMERASE; MEMBRANE.
CC KW ISOMERASE; 658 AA; 74068 MW; 2CBB058A CRC32;
CC SQ SEQUENCE 658 AA; 74068 MW; 2CBB058A CRC32;

Query Match 61.0%; Score 36; DB 1; Length 658;
Best Local Similarity 63.6%; Pred. No. 47;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LKPEQILQLL 12
I::I::I::I::
Db 355 LKPEQILQLL 365

RESULT 12
TIR_ECOLI
ID TIR-ECOLI STANDARD; PRT; 1188 AA.
AC P08956;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYPE I RESTRICTION ENZYME ECKO I R PROTEIN (EC 3.1.21.3).
GN HSDR OR HSR.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 88118919.
RA LOENEN W.A.M., DANIEL A.S., BRAYMER H.D., MURRAY N.E.;
RT "Organization and sequence of the hsd genes of Escherichia coli
RT K-12.";
RL J. MOL. BIOL. 198:159-170(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 95334362.
RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
RA BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL NUCLEIC ACIDS RES. 23:2105-2119(1995).
RN [3]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN-K12 / CR63;
RX MEDLINE; 91317743.
RA WAITE-REES P.A., KEATING C.J., MORAN L.S., SLATKO B.E., HORNSTRA L.J.,
RA BENNER J.S.;
RT "Characterization and expression of the Escherichia coli Mrr
RT restriction system.";
RL J. BACTERIOL. 173:5207-5219(1991).
CC -!- FUNCTION: THE ECKO I ENZYME RECOGNIZES 5'AAC(N6)GTGC.
CC -!- FUNCTION: SUBUNIT R IS REQUIRED FOR BOTH NUCLEASE AND ATPASE
CC ACTIVITIES, BUT NOT FOR MODIFICATION.
CC -!- SUBUNIT: THE TYPE I RESTRICTION & MODIFICATION SYSTEM IS COMPOSED
CC OF THREE POLYPEPTIDES R, M AND S.
CC -!- TYPE I RESTRICTION AND MODIFICATION ENZYMES ARE COMPLEX, MULTI-
CC FUNCTIONAL SYSTEMS WHICH REQUIRE ATP. S-ADENOSYL METHIONINE AND
CC MG(2+) AS CO-FACTORS AND, IN ADDITION TO THEIR ENDONUCLEOLYTIC
CC AND METHYLASE ACTIVITIES, ARE POTENT DNA-DEPENDENT ATPASES.
CC -!- SIMILARITY: WITH ATPASES.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1084

CC ONWARD AND IS SHORTER (1090 AA) DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X06545; G41752; ALT_FRAME.
CC DR EMBL: U14003; G537192; -.
CC DR EMBL: AE000505; G1790809; -.
CC DR EMBL: X54198; G42019; ALT_INIT.
CC DR PIR: Q00648; NDECKR.
CC DR REBASE: RB00001; ECOKI.
CC DR ECGENE: EG10459; HSDR.
CC DR PFAM: PF00271; helicase.C; 1.
CC KW RESTRICTION SYSTEM; HYDROLASE; DNA-BINDING; ATP-BINDING.
CC FT DNA_BIND 449 468 H-T-H MOTIF (BY SIMILARITY).
CC FT NP_BIND 482 520 ATP (BY SIMILARITY).
CC FT CONFLICT 629 697 DAVKIALTATPALHTVQIFGEPIRYTYRTAVIDGLIDQD
CC PPIQITRNAOEGVLSKGEQVERISQ -> ECGNRSR
CC HPGATYCADFRRAGLPLYPYRGYRFRSDRPGSAYSDHRPQ
CC RAGGGLSLQRRAGRAHQ (IN REF. 1).
CC SQ SEQUENCE 1188 AA; 136100 MW; D9E3EA2C CRC32;

Query Match 61.0%; Score 36; DB 1; Length 1188;
Best Local Similarity 50.0%; Pred. No. 90;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 3 KPEQILQLLG 12
:||:|:|:
Db 433 RPEELMLG 442

RESULT 13
TALA_DICDI STANDARD; PRT; 2491 AA.
ID TALA_DICDI
AC P54633;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE FILOPODIN (TALIN HOMOLOG).
GN TALA.
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AX2;
RX MEDLINE: 95213284.
RA KREITMEIER M., GERISCH G., HEIZER C., MUELLER-TAUBENBERGER A.;
RT "A talin homolog of Dictyostelium rapidly assembles at the leading
RT edge of cells in response to chemoattractant.";
RL J. CELL BIOL. 129:179-188(1995).
CC -!- FUNCTION: ACTIN-BINDING PROTEIN THAT MAY BE INVOLVED IN THE
CC CONTROL OF CELL MOTILITY AND CHEMOTAXIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC; RAPIDLY ASSEMBLES AT THE
CC LEADING EDGE OF CELLS IN RESPONSE TO CHEMOATTRACTANT.
CC -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,
CC RADIXIN, AND TALIN.
CC -!- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C.ELEGANS ZK370.3.
CC PARTIAL, TO YEAST ROD1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: U14576; G797299; -.
DR DICTYDS; DD01121; TALA.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00661; BAND_41_3; 1.
DR PFAM; PF00373; Band_41; 1.
KW STRUCTURAL PROTEIN; CYTOSKELETON; ACTIN-BINDING.
FT DOMAIN 134 294 BAND 4.1-LIKE DOMAIN.
SQ SEQUENCE 2491 AA; 268811 MW; C534F8BB CRC32;

Query Match 61.0%; Score 36; DB 1; Length 2491;
Best Local Similarity 77.8%; Pred. No. 2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 PEQILQLLG 12
:|:|:|:|:
Db 347 PEQISQLIG 355

RESULT 14
ZNUC_ECOLI STANDARD; PRT; 251 AA.
ID ZNUC_ECOLI
AC P52848; P76285;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE HIGH-AFFINITY ZINC UPTAKE SYSTEM ATP-BINDING PROTEIN ZNUC.
GN ZNUC.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / EMG2;
RA ROBISON K., O'KEEFE T., CHURCH G.M.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL SCIENCE 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE: 97251358.
RA ITOH T., AIBA H., BABA T., FUJITA K., HAYASHI K., INADA T.,
RA ISONO K., KASAI H., KIMURA S., KITAKAWA M., KITAGAWA M.,
RA MAKINO K., MIKI T., MIZOBUCHI K., MORI H., MORI T., MOTOMURA K.,
RA NAKADE S., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,
RA SAITO N., SAMEI G., SEKI Y., SIVASUNDARAM S., TAGAMI H.,
RA TAKEDA J., TAKEMOTO K., WADA C., YAMAMOTO Y., HORIUCHI T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA RES. 3:379-392(1996).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE: 98343803.
RA PATZER S.I., HANTKE K.;
RT "The znABC high-affinity zinc uptake system and its regulator Zur in
RT Escherichia coli.";
RL MOL. MICROBIOL. 28:1199-1210(1998).
CC -!- FUNCTION: INVOLVED IN THE HIGH-AFFINITY ZINC UPTAKE TRANSPORT
CC SYSTEM.
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). STRONG, TO H-INFLUENZAE ZNUC.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 188.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U38702; NOT_ANNOTATED_CDS.
DR EMBL; AE000280; GI788165; -.
DR EMBL; D90828; GI736501; -.
DR EMBL; D90829; GI736505; -.
DR EMBL; EG13132; ZNUC.
DR PROSITE; PS00211; ABC_TRANSPORTER; FALSE_NEG.
DR PFAM; PF00005; ABC_tran; 1.
KW ATP-BINDING; TRANSPORT; ZINC.
FT NP_BIND 37 44 ATP (POTENTIAL).
FT CONFLICT 80 80 T -> S (IN REF. 1).
FT CONFLICT 83 83 P -> S (IN REF. 1).
FT CONFLICT 135 135 L -> W (IN REF. 1).
FT CONFLICT 143 143 V -> A (IN REF. 1).
FT CONFLICT 155 155 G -> W (IN REF. 1).
FT CONFLICT 174 174 V -> A (IN REF. 1).
FT CONFLICT 176 176 M -> L (IN REF. 1).
SQ SEQUENCE 251 AA; 27867 MW; B81520FC CRC32;

Query Match 51.0%; Score 36; DB 1; Length 251;
Best Local Similarity 72.7%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 LKPEQILQLLG 12
|||:|||||
Db 27 LKPKILTLG 37

RESULT 15

ARP7_YEAST
ID ARP7_YEAST STANDARD; PRT; 477 AA.
AC Q12406;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ACTIN-LIKE PROTEIN ARP7.
GN ARP7 OR YP034W OR YP9367.14.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA BADCOCK K., BOWMAN S., CHURCHER C.M., PEARSON D., RAJANDREAM M.A.,
RA WALSH S.V., BARRELL B.G.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP GENE NAME.
RX MEDLINE; 97435478.
RA POCH O., WINSOR B.;
RT "Who's who among the Saccharomyces cerevisiae actin-related proteins?
RT A classification and nomenclature proposal for a large family.";
RL YEAST 13:1053-1058(1997).
CC -!- SIMILARITY: TO ACTINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; Z71255; E236895; -.

DR EMBL; Z49274; G809599; -.
DR SGD; L0003436; ARP7.
DR PFAM; PF00022; actin; 1.
KW STRUCTURAL PROTEIN; CYTOSKELETON.
SQ SEQUENCE 477 AA; 53810 MW; CC82222CA CRC32;

Query Match 59.3%; Score 35; DB 1; Length 477;
Best Local Similarity 54.5%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NLKPEQILQLL 11
|: |||: ||
Db 378 NISPEQVVISLL 388

Search completed: November 13, 1999, 10:33:46
Job time: 5196 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:55:40 ; Search time 139.86 Seconds
(without alignments)
5.280 Million cell updates/sec

Title: US-08-913-430-14
Perfect score: 59
Sequence: 1 NLKPEQILQLLG 12

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database : SPTRMBL_10:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	89.8	627	2	O30643 mycoplasma
2	53	89.8	625	2	O30704 mycoplasma
3	41	69.5	1203	4	O95341 homo sapien
4	41	69.5	1146	4	O95856 homo sapien
5	38	64.4	1182	11	O35595 mus musculus
6	38	64.4	778	12	O84509 paramesium
7	37	62.7	897	4	O92814 homo sapien
8	37	62.7	528	12	O89925 african swi
9	36	61.0	589	2	O50642 porphyromon
10	36	61.0	283	2	O51941 leptospira
11	36	61.0	392	2	O44305 acinetobact
12	36	61.0	735	2	P94180 anabaena sp
13	36	61.0	368	2	P74127 synechocyst
14	36	61.0	283	2	O92HF9 leptospira
15	36	61.0	1070	3	P78734 pneumocysti
16	36	61.0	453	3	O86671 saccharomyc
17	36	61.0	3134	5	O25994 plasmodium
18	36	61.0	562	5	O17330 caenorhabdi
19	36	61.0	222	5	O23274 caenorhabdi
20	36	61.0	623	5	O15873 plasmodium
21	36	61.0	1244	5	O21778 caenorhabdi
22	36	61.0	704	5	O93376 caenorhabdi
23	36	61.0	729	11	O35095 rattus norv
24	36	61.0	618	11	O88985 mus musculus
25	36	61.0	729	11	O920E0 mus musculus
26	35	60.2	1522	2	O92ID9 actinobacil
27	35	59.3	446	2	O07890 treponema p
28	35	59.3	447	2	O26027 helicobacte
29	35	59.3	447	2	O92JC6 helicobacte

30	35	59.3	706	3	Q12039 saccharomyc
31	35	59.3	744	4	O15404 homo sapien
32	35	59.3	559	4	O99799 homo sapien
33	35	59.3	718	5	O77136 apis mellif
34	35	59.3	688	8	Q34312 dictyosteli
35	35	59.3	1259	10	O40983 pismum savi
36	35	59.3	353	10	O80449 arabidopsis
37	35	59.3	293	11	O35447 mus musculu
38	35	59.3	618	11	O88984 rattus norv
39	35	59.3	475	12	O84501 paramesium
40	35	59.3	1083	12	O39483 avian sarco
41	35	59.3	381	12	O9YMW5 lymantria d
42	35	59.3	557	13	O90999 gallus gall
43	35	59.3	358	13	O91758 xenopus lae
44	35	59.3	1068	13	O42391 gallus gall
45	34	57.6	585	13	Q9YH33 oreochromis

ALIGNMENTS

RESULT 1

O30643 PRELIMINARY; PRT; 627 AA.

AC O30643;

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)

DE PROLIPOPROTEIN P65 PRECURSOR.

OS Mycoplasma hyopneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Mycoplasmataceae; Mycoplasma.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-J, ATCC25934;

RA HEIDARI M.B., KIM M.F., MCINTOSH M.A., WISE K.W.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF013714; AAB67173.1; -

DR PFAM; PF00657; Lipase_GDSL; 1.

KW Signal; Lipoprotein..

FT SIGNAL

FT CHAIN

SQ SEQUENCE 627 AA; 71016 MW; 10C5A2A5 CRC32;

Query Match 89.8%; Score 53; DB 2; Length 627;
Best Local Similarity 91.7%; Pred. No. 0.16;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
|||||

DB 551 NLKPEQILQLLG 562

RESULT 2

O30704 PRELIMINARY; PRT; 625 AA.

AC O30704;

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)

DE SURFACE LIPOPROTEIN PRECURSOR.

OS Mycoplasma hyopneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Mycoplasmataceae; Mycoplasma.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-232A;

RA HSU T., MINION F.C.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF015665; AAB70214.1; -

DR PFAM; PF00657; Lipase_GDSL; 1.

KW Signal; Lipoprotein.

FT SIGNAL 1 31 POTENTIAL.
 FT CHAIN 32 625 MATURE SURFACE LIPOPROTEIN.
 SQ SEQUENCE 625 AA; 70914 MW; C109763D CRC32;

Query Match 89.8%; Score 53; DB 2; Length 625;
 Best Local Similarity 91.7%; Pred. No. 0.16;
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
 |||||
 Db 549 NLKPEQILTLG 560

RESULT 3
 O95341 PRELIMINARY; PRT; 1203 AA.
 AC O95341;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE RECEPTOR PROTEIN PATCHED 2.
 GN PTCH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99030620.
 RA CARPENTER D., STONE D.M., BRUSH J., RYAN A., ARMANINI M., FRANTZ G.,
 RA ROSENTHAL A., DE SAUVAGE F.J.;
 RT "Characterization of two patched receptors for the vertebrate
 RT hedgehog protein family";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:13630-13634(1998).
 DR EMBL; AF091501; AAC79847.1; -.
 KW Receptor.
 SQ SEQUENCE 1203 AA; 130573 MW; 2182A764 CRC32;

Query Match 69.5%; Score 41; DB 4; Length 1203;
 Best Local Similarity 66.7%; Pred. No. 41;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLKPEQILQLLG 12
 |||||
 Db 215 NLDPEQLLEELG 226

RESULT 4
 O95856 PRELIMINARY; PRT; 1146 AA.
 AC O95856;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE PATCHED 2.
 GN PTCH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA ZAPHIROPOULOS P.G., UNDEN A.B., RAHNAMA F., HOLLINGSWORTH R.E.,
 RA TOFTGARD R.;
 RT "PTCH2, a novel human patched gene, undergoing alternative splicing
 RT and upregulated in basal cell carcinomas";
 RL Cancer Res. 0:0-0(1999).
 DR EMBL; AF119569; AAD17260.1; -.
 SQ SEQUENCE 1146 AA; 124958 MW; ECE4AC40 CRC32;

Query Match 69.5%; Score 41; DB 4; Length 1146;
 Best Local Similarity 66.7%; Pred. No. 39;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 NLKPEQILQLLG 12
 |||||
 Db 215 NLDPEQLLEELG 226

RESULT 5
 O35595 PRELIMINARY; PRT; 1182 AA.
 AC O35595;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE PATCHED PROTEIN HOMOLOG 2 (PTC2).
 GN PTCH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-EMBRYO;
 RX MEDLINE; 98122566.
 RA MOTOTAMA J., TAKABATAKE T., TAKESHIMA K., HUI C.-C.;
 RT "ptch2, a second mouse patched gene is co-expressed with Sonic
 RT hedgehog";
 RL Nat. Genet. 18:104-106(1998).
 RN [2]
 RP SEQUENCE OF 196-446 FROM N.A.
 RC STRAIN-BALB/C; TISSUE-NEURETINA;
 RX MEDLINE; 97379366.
 RA TAKABATAKE T., OGAWA M., TAKAHASHI T.C., MIZUNO M., OKAMOTO M.,
 RA TAKESHIMA K.;
 RT "hedgehog and patched gene expression in adult ocular tissues";
 RL FEBS Lett. 410:485-489(1997).
 CC -!- FUNCTION: MAY HAVE A ROLE IN EPIDERMAL DEVELOPMENT. MAY ACT AS A
 CC RECEPTOR FOR THE SONIC HEDGEHOG PROTEIN (SHH).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN EPITHELIAL CELLS OF THE
 CC DEVELOPING HAIR, TOOTH AND WHISKER.
 CC -!- DEVELOPMENTAL STAGE: DETECTED IN 8.5 TO 17.5 DAYS POST COITUM
 CC EMBRYOS.
 CC -!- SIMILARITY: BELONGS TO PATCHED FAMILY.
 CC EMBL; AB010833; BAA24691.1; -.
 DR MGD; MGI:1095405; PTCH2.
 KW Transmembrane; Developmental
 FT DOMAIN 1 57
 FT TRANSMEM 58 78
 FT DOMAIN 79 394
 FT TRANSMEM 395 414
 FT DOMAIN 415 428
 FT TRANSMEM 429 449
 FT DOMAIN 450 457
 FT TRANSMEM 458 478
 FT DOMAIN 479 501
 FT TRANSMEM 502 522
 FT DOMAIN 523 531
 FT TRANSMEM 532 552
 FT DOMAIN 553 686
 FT TRANSMEM 687 707
 FT DOMAIN 708 963
 FT TRANSMEM 964 984
 FT DOMAIN 985 991
 FT TRANSMEM 992 1012
 FT DOMAIN 1013 1013
 FT TRANSMEM 1014 1034
 FT DOMAIN 1035 1064
 FT TRANSMEM 1065 1085
 FT DOMAIN 1086 1086
 FT TRANSMEM 1087 1107
 FT DOMAIN 1108 1182
 FT CARBOHYD 370
 FT CARBOHYD 812

Query Match 62.7%; Score 37; DB 4; Length 897;
Best Local Similarity 63.6%; Pred. No. 1.6e+02;
Matches 7; Conservative 2; Mismatches 2; Indels

RC STRAIN-BA71V;
RX MEDLINE; 93346971.
RA YANEZ R.J., RODRIGUEZ J.M., RODRIGUEZ J.F., SALAS M.L., VINUELA E.;
RT "African swine fever virus thymidylate kinase gene: sequence and
transcriptional mapping."
RL J. Gen. Virol. 74:1633-1638(1993).
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 94065656.
RA ALCAMI A., ANGULO A., VINUELA E.;
RT "Mapping and sequence of the gene encoding the African swine fever
virus protein of M(r) 11500."
RL J. Gen. Virol. 74:2317-2324(1993).
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 93277388.
RA MUNOZ M., FREIJE J.M., SALAS M.L., VINUELA E., LOPEZ-OTIN C.;
RT "Structure and expression in E. coli of the gene coding for protein
p10 of African swine fever virus."
RL Arch. Virol. 130:93-107(1993).
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 90357780.
RA BLASCO R., LOPEZ-OTIN C., MUNOZ M., BOCKAMP E.O., SIMON-MATEO C.,
VINUELA E.;
RT "Sequence and evolutionary relationships of African swine fever virus
thymidine kinase."
RL Virology 178:301-304(1990).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 93281390.
RA YANEZ R.J., BOURSNEILL M., NOGAL M.L., YUSTE L., VINUELA E.;
RT "African swine fever virus encodes two genes which share significant
homology with the two largest subunits of DNA-dependent RNA
polymerases."
RL Nucleic Acids Res. 21:2423-2427(1993).
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 93353606.
RA RODRIGUEZ J.M., YANEZ R.J., ALMAZAN F., VINUELA E., RODRIGUEZ J.F.;
RT "African swine fever virus encodes a CD2 homolog responsible for the
adhesion of erythrocytes to infected cells."
RL J. Virol. 67:5312-5320(1993).
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 94085774.
RA YANEZ R.J., RODRIGUEZ J.M., BOURSNEILL M., RODRIGUEZ J.F., VINUELA E.;
RT "Two putative African swine fever virus helicases similar to yeast
'DEAH' pre-mRNA processing proteins and vaccinia virus ATPases D1L1
and D6R."
RL Gene 134:161-174(1993).
RN [15]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 90223993.
RA LOPEZ-OTIN C., FREIJE J.M., PARRA F., MENDEZ E., VINUELA E.;
RT "Mapping and sequence of the gene coding for protein p72, the major
capsid protein of African swine fever virus."
RL Virology 175:477-484(1990).
RN [16]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 94123986.
RA RODRIGUEZ J.M., YANEZ R.J., RODRIGUEZ J.F., VINUELA E., SALAS M.L.;
RT "The DNA polymerase-encoding gene of African swine fever virus:
sequence and transcriptional analysis."
RL Gene 136:103-110(1993).

RN [17]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 93327788.
RA SIMON-MATEO C., ANDRES G., VINUELA E.;
RT "Polyprotein processing in African swine fever virus: a novel gene
expression strategy for a DNA virus."
RL EMBO J. 12:2977-2987(1993).
RN [18]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 93233210.
RA PRADOS F.J., VINUELA E., ALCAMI A.;
RT "Sequence and characterization of the major early phosphoprotein p32
of African swine fever virus."
RL J. Virol. 67:2475-2485(1993).
RN [19]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 92260660.
RA ALCAMI A., ANGULO A., LOPEZ-OTIN C., MUNOZ M., FREIJE J.M.,
CARRASCOA A.L., VINUELA E.;
RT "Amino acid sequence and structural properties of protein p12, an
African swine fever virus attachment protein."
RL J. Virol. 66:3860-3868(1992).
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 93174976.
RA YANEZ R.J., VINUELA E.;
RT "African swine fever virus encodes a DNA ligase."
RL Virology 193:531-536(1993).
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 93174941.
RA PENA L., YANEZ R.J., REVILLA Y., VINUELA E., SALAS M.L.;
RT "African swine fever virus guanylyltransferase."
RL Virology 193:319-328(1993).
RN [22]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 95159428.
RA SIMON-MATEO C., FREIJE J.M., ANDRES G., LOPEZ-OTIN C., VINUELA E.;
RT "Mapping and sequence of the gene encoding protein p17, a major
African swine fever virus structural protein."
RL Virology 206:1140-1144(1995).
RN [23]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 92263807.
RA GARCIA-BEATO R., FREIJE J.M., LOPEZ-OTIN C., BLASCO R., VINUELA E.,
SALAS M.L.;
RT "A gene homologous to topoisomerase II in African swine fever
virus."
RL Virology 188:938-947(1992).
RN [24]
RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE; 94091056.
RA FREIJE J.M., LAIN S., VINUELA E., LOPEZ-OTIN C.;
RT "Nucleotide sequence of a nucleoside triphosphate phosphohydrolase
Query Match 62.7%; Score 37; DB 12; Length 528;
Best Local Similarity 70.0%; Pred. No. 91;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLKPEQIQL 10

Db 439 SLKPEILXL 448

RESULT 9

O50642
ID O50642 PRELIMINARY; PRT; 689 AA.
AC O50642;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PEPO.
GN PEPO.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Cytophagales; Bacteroidaceae; Porphyromonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-381;
RA ANSAI T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB010440; BAA24495.1; -.
DR PFAM; PF01431; Peptidase_M13; 1.
SQ SEQUENCE 689 AA; 78845 MW; A3F83E95 CRC32;
[1]
Query Match 61.0%; Score 36; DB 2; Length 689;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
QY 1 NLKPEQILQL 10
Db 634 NIRPEVLRL 643
[1]:::1:1
RESULT 10
ID O51941 PRELIMINARY; PRT; 283 AA.
AC O51941;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PERIPLASMIC FLAGELLIN.
GN FLAB.
OS Leptospira interrogans.
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-POMONA;
RA LIN M., SURUBALLI O., NIELSEN K., NADIN-DAVIS S., RANDALL G.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF014114; AAB94024.1; -.
DR PFAM; PF00700; Flagellin_C; 1.
DR PFAM; PF00669; Flagellin_N; 1.
SQ SEQUENCE 283 AA; 31294 MW; 848EDA89 CRC32;
[1]
Query Match 61.0%; Score 36; DB 2; Length 283;
Best Local Similarity 54.5%; Pred. No. 73;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 NLKPEQILQL 11
Db 270 NMRPNSVLKLL 280
[1]:::1:1
RESULT 11
ID Q44305 PRELIMINARY; PRT; 392 AA.
AC Q44305;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE BETA-KETOTHIOLASE.
GN PHAA.
OS Acinetobacter sp.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Moraxellaceae; Acinetobacter.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-RA3849;
RX MEDLINE; 95362679.
RA SCHEMBRI M.A., BAYLY R.C., DAVIES J.K.;
RT "phosphate concentration regulates transcription of the Acinetobacter
RT polyhydroxyalkanoic acid biosynthetic genes.";
RL J. Bacteriol. 177:4501-4507(1995).
DR EMBL; L37761; AAA99475.1; -.
DR PFAM; PF00108; thiolase; 1.
SQ SEQUENCE 392 AA; 40630 MW; 3AA0165C CRC32;
[1]
Query Match 61.0%; Score 36; DB 2; Length 392;
Best Local Similarity 54.5%; Pred. No. 1e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 NLKPEQILQL 11
Db 42 NVKPEQDVVI 52
[1]:::1:1
RESULT 12
ID P94180 PRELIMINARY; PRT; 735 AA.
AC P94180;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE ADENYLATE CYCLASE (EC 4.6.1.1).
GN CYAA.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC7120;
RX MEDLINE; 97315230.
RA KATAYAMA M., OHMORI M.;
RT "Isolation and characterization of multiple adenylate cyclase genes
RT from the cyanobacterium Anabaena sp. strain PCC 7120.";
RL J. Bacteriol. 179:3588-3593(1997).
DR EMBL; D89622; BAA13997.1; -.
DR PFAM; PF00211; guanylate_cyc; 1.
KW Lyase.
SQ SEQUENCE 735 AA; 80755 MW; E1668B66 CRC32;
[1]
Query Match 61.0%; Score 36; DB 2; Length 735;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 LKPEQILQL 11
Db 459 LSPQVQVQL 468
[1]:::1:1
RESULT 13
ID P74127 PRELIMINARY; PRT; 368 AA.
AC P74127;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ABC TRANSPORTER.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SQ SEQUENCE FROM N.A.
RC STRAIN-PCC6803;

RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.,
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90912; BAAL8213.1; -.
DR PFAM; PF00005; ABC_tran; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 368 AA; 40620 MW; 8E39B600 CRC32;

Query Match 61.0%; Score 36; DB 2; Length 368;
Best Local Similarity 66.7%; Pred. No. 95;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NLKPEQILQLLG 12
||| :|||
Db 42 NLAPGEILGLG 53

RESULT 14
Q9ZHF9 PRELIMINARY; PRT; 283 AA.
AC Q9ZHF9;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE FLAGELLIN.
GN FLAB.
OS Leptospira borgpetersenii.
OC Bacteria; Spirochaetales; Leptospiraceae; Leptospira.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FIELD STRAIN;
RA LIN M.;
RT "Molecular analysis of flab, a periplasmic flagellar core protein
RT gene in pathogenic leptospires.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF064057; AAC99336.1; -.
SQ SEQUENCE 283 AA; 31324 MW; 888105D5 CRC32;

Query Match 61.0%; Score 36; DB 2; Length 283;
Best Local Similarity 54.5%; Pred. No. 73;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NLKPEQILQLL 11
||| :|||
Db 270 NKPNSVLKLL 280

RESULT 15
P78734 PRELIMINARY; PRT; 1070 AA.
AC P78734;
DT 01-MAY-1997 (TRENBLrel. 03, Created)
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DT 01-AUG-1998 (TRENBLrel. 07, Last annotation update)
DE MAJOR SURFACE GLYCOPROTEIN.
OS Pneumocystis carinii.
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes; Pneumocystidaceae;
OC Pneumocystis.
RN [1]
RP SEQUENCE FROM N.A.
RA EDMAN J.C., HATTON T.W., KOVACS J.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U83323; ABA41323.1; -.

SQ SEQUENCE 1070 AA; 120810 MW; 7B3FD4DE CRC32;
Query Match 61.0%; Score 36; DB 3; Length 1070;
Best Local Similarity 58.3%; Pred. No. 2.8e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NLKPEQILQLLG 12
||| :|||
Db 621 NIKNEILQEIG 632

Search completed: November 13, 1999, 12:55:42
Job time: 3041 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 19:00:03 ; Search time 104.22 Seconds
(without alignments)
4.318 Million cell updates/sec

Title: US-08-913-430-15
Perfect score: 81
Sequence: 1 LKAXNKXIEINTXLDN 19

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	92.6	19	1 W01027	Mycoplasma 72-75 k
2	75	92.6	627	1 W62451	Mycoplasma hyopneu
3	40	49.4	846	1 R08406	Sequence deduced f
4	39	48.1	639	1 R47172	Sequence of the in
5	38	46.9	2366	1 R95011	C. difficile toxin
6	38	46.9	2366	1 W63888	Clostridium diffic
7	37	45.7	147	1 W18514	Human RC-9 implica
8	37	45.7	669	1 W56486	Zea mays soluble s
9	37	45.7	147	1 W85471	ATG-750/RC-9 (allo
10	37	45.7	147	1 W85469	ATG-750/RC-9 (allo
11	37	45.7	133	1 W85470	ATG-1120 (allograf
12	36	44.4	275	1 R43390	Sequence specific
13	36	44.4	337	1 R63881	Meropsin amino acid
14	36	44.4	2482	1 R72826	Human mitotin. Pur
15	36	44.4	3110	1 R71730	Meropsin major subu
16	36	44.4	337	1 R94376	Humam meropsin par
17	36	44.4	275	1 R98993	Neisseria gonorrhoe
18	36	44.4	3248	1 R99795	Kinetochore protei
19	36	44.4	337	1 W39206	Human partial mero
20	36	44.4	2482	1 W23996	Human mitotin amin
21	36	44.4	311	1 W98846	H. pylori GHPO 159
22	36	44.4	176	1 Y00016	Enterococcus faeca
23	36	44.4	156	1 Y00017	Enterococcus faeca
24	35	43.2	581	1 R42384	H. somnus lppC. Ha
25	35	43.2	33	1 W45306	Glucagon-like pept
26	35	43.2	33	1 W45275	Glucagon-like pept
27	35	43.2	33	1 W45288	Glucagon-like pept
28	35	43.2	157	1 W41151	RBE1 transcripion
29	35	43.2	445	1 W71558	Helicobacter polyp
30	35	43.2	506	1 W98863	H. pylori GHPO 167
31	35	43.2	445	1 W98793	H. pylori GHPO 125
32	35	43.2	1279	1 W98761	H. pylori GHPO 109
33	35	43.2	1048	1 W97720	Staphylococcus aur
34	34	42.0	573	1 R04713	Amino acid sequenc
35	34	42.0	1784	1 R05898	Gene product of fi
36	34	42.0	195	1 P60400	Equine IFN-omega-1
37	34	42.0	1116	1 R12083	HWP protein. New p
38	34	42.0	248	1 R12532	Human Macrophage C
39	34	42.0	870	1 R14529	Mosquitotoxin. Cau
40	34	42.0	870	1 R41020	Insecticidal prote
41	34	42.0	524	1 R41681	Reverse transcript
42	34	42.0	547	1 R67385	Mitochondrial prot
43	34	42.0	575	1 R64763	Human P1 protein.

ALIGNMENTS

44 34 42.0 1045 1 R66221 Sucrose-phosphate-
45 34 42.0 33 1 W78290 Fragment of human

RESULT 1

W01027
ID W01027 standard; Peptide; 19 AA.
AC W01027:
DT 19-JAN-1997 (first entry)
DE Mycoplasma 72-75 kDa protective antigen peptide CNBR F3.
KW Antigen; vaccine; mycoplasma pneumoniae; swine enzootic pneumonia;
diagnosis; antibody.
OS Mycoplasma hyopneumoniae strain Beaufort.
FH Key Location/Qualifiers
FT misc_difference 6
FT misc_difference 9 /note= "undetermined amino acid"
FT misc_difference 16 /note= "undetermined amino acid"
FT misc_difference 16 /note= "undetermined amino acid"
PN W09628472-Al.
PD 19-SEP-1996.
PF 15-MAR-1996; AU0149.
PR 16-MAR-1995; AU-001789.
PA (UYME) UNIV MELBOURNE.
PI Doughty SW Lee R, Walker J;
DR WPI; 96-433763/43.
PT Putative protective antigens against Mycoplasma - used for the
detection, prevention or treatment of Mycoplasma infections, esp. M.
hyopneumoniae in swine
PS Claim 8: Page 27: 43pp: English.
CC A 72-75 kDa putative protective antigen against Mycoplasma
contains the N-terminal sequence given in W01024 and the internal
CNBR fragments given in W01025-27. The antigen was isolated from
Mycoplasma hyopneumoniae cells using antibody probes enriched with
Mycoplasma-specific antibodies. Other protective antigens were
also identified (see also W01028-37). Protective antigens and
antibodies can be used in vaccines for preventing or treating
mycoplasma infections, partic. M. hyopneumoniae infections in
swine. They can also be used for diagnosis.
SQ Sequence 19 AA;

Query Match 92.6%; Score 75; DB 1; Length 19;
Best Local Similarity 100.0%; Pred. No. 5.9e-07;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKAXNKXIEINTXLDN 19

|||||
Db 1 LKAXNKXIEINTXLDN 19

RESULT 2

W62451
ID W62451 standard; Protein; 627 AA.
AC W62451;
DT 01-OCT-1998 (first entry)
DE Mycoplasma hyopneumoniae P65 surface antigen.
KW Mycoplasma hyopneumoniae; surface lipoprotein P65; structural gene;
antigen; vaccine; swine; enzootic pneumoniae; diagnosis; pig;
immune response; mycoplasma pneumoniae.
OS Mycoplasma hyopneumoniae.
PN US5788962-A.
PD 04-AUG-1998.
PF 28-AUG-1996; 703947.
PR 17-JAN-1995; US-373957.
PR 28-AUG-1996; US-703947.
PA (UMOR) UNIV MISSOURI.
PI McIntosh MA, Wise KS;
DR WPI; 98-446004/38.


```

PT RC-9 polynucleotide(s), proteins and antibodies, e.g. to identify
PT inhibitors of RC-9 activity
PS Claim 6; Page 23-24; 34pp; English.
CC RC-9 is a protein implicated in coronary arterial disease. The
CC human RC-9 amino acid sequence (W18514) was deduced from a cDNA
CC clone (T64809). Tissue analysis has demonstrated elevated RC-9
CC mRNA in association with hypertension, congestive heart failure,
CC and myocardial infarction consistent with a role of human RC-9
CC in the control of cellular proliferation/differentiation and
CC the pathogenesis of fibroproliferative disorders. RC-9
CC polypeptides can be expressed in host cells and used in methods of
CC diagnosing proliferative arterial disease, to produce antibodies
CC useful in treating such disease, and to screen for (ant)agonists.
CC Anti-RC-9 antibodies may be used to treat disorders characterised
CC by vascular restenosis, e.g. atherosclerosis.
SQ Sequence 147 AA;

Query Match 45.7%; Score 37; DB 1; Length 147;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LKAXNKXIEINTXLD 14
   | | | | : : | | |
Db 15 LKAAQOEERLDEIN 28

RESULT 8
W56486
ID W56486 standard; Protein; 669 AA.
AC W56486;
DT 11-SEP-1998 (first entry)
DE Zea mays soluble starch synthase IIA.
KW SER; starch-encapsulating region; fusion vector;
KW soluble starch synthase IIA; glucosyl transferase.
OS Zea mays.
PN W09814601-Al.
PD 09-APR-1998.
PF 30-SEP-1997; U17555.
PR 30-SEP-1996; US-026855.
PA (EXSE-) EXSEED GENERICS LLC.
PI Guan H, Keeling P;
PI WPI; 98-240100/21.
DR N-PSDB; V29754#.
PT Hybrid polypeptide comprising starch-encapsulating region and
PT protein - useful for, e.g. producing protein(s) resistant to
PT degradation by stomach acids
PS Example 2; Page 34-36; 156pp; English.
CC The sequence is that of soluble starch synthase IIA.
CC It can be used in the production of a hybrid polypeptide
CC comprising a starch-encapsulating region (SER) fused
CC to a payload protein. The hybrid polypeptide can be used to make
CC modified starches comprising the payload protein, selected from,
CC e.g. hormones, growth factors, antibodies, enzymes, dyes,
CC immunoglobulins, etc. The modified starch can also be used
CC to provide grain feeds enriched in amino acids. By encapsulating
CC the payload protein in starch, it is more resistant to
CC degradation by stomach acids.
SQ Sequence 669 AA;

Query Match 45.7%; Score 37; DB 1; Length 669;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 KAXXNKXIEINTXLD 18
   : | | | | | : |
Db 615 RAAANKLIEVLSHCLD 630

RESULT 9
W85471
ID W85471 standard; Protein; 147 AA.

```

```

Query Match 46.9%; Score 38; DB 1; Length 2366;
Best Local Similarity 41.2%; Pred. No. 2.9e+02;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 LKAXNKXIEINTXLD 18
   | | | | : : | | |
Db 210 LSNEYSKEIDELNTYIE 226

RESULT 6
W88388
ID W88388 standard; Protein; 2366 AA.
AC W88388;
DT 07-DEC-1998 (first entry)
DE Clostridium difficile toxin B.
KW Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;
KW Pseudomonas enterocolitis.
OS Clostridium difficile.
PN W09808540-Al.
PD 05-MAR-1998.
PF 28-AUG-1997; U15394.
PR 28-AUG-1996; US-704159.
PA (OPHI-) OPHIDIAN PHARM INC.
PI Thalley BS, Williams JA;
PI WPI; 98-230234/20.
DR N-PSDB; V30561.
PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
PS Example 18; Page 241-249; 428pp; English.
CC This is the amino acid sequence of Clostridium difficile toxin B,
CC deduced from the coding region (see V30561) of the toxin B gene.
CC Fragments of the toxin B gene have been cloned into various
CC prokaryotic expression systems, and assessed for the ability to
CC express recombinant toxin B protein in E. coli. It would be
CC advantageous to use simple and inexpensive prokaryotic expression
CC systems to produce and purify high levels of recombinant toxin B
CC for immunisation purposes. The invention specifically relates to
CC recombinant proteins derived from Clostridium botulinum toxins
CC (see W68389-400) and their use as immunogens for the production of
CC vaccines and antitoxins.
SQ Sequence 2366 AA;

Query Match 46.9%; Score 38; DB 1; Length 2366;
Best Local Similarity 41.2%; Pred. No. 2.9e+02;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 LKAXNKXIEINTXLD 18
   | | | | : : | | |
Db 210 LSNEYSKEIDELNTYIE 226

RESULT 7
W18514
ID W18514 standard; Protein; 147 AA.
AC W18514;
DT 03-SEP-1997 (first entry)
DE Human RC-9 implicated in coronary arterial disease.
KW RC-9; proliferative arterial disease; vascular restenosis;
KW diagnosis; therapy; antibody; atherosclerosis.
OS Homo sapiens.
PN W09722860-Al.
PD 26-JUN-1997.
PF 13-DEC-1996; U19671.
PR 18-DEC-1995; US-008801.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Arleth AJ, Autieri MV, Ohlstein EH;
PI WPI; 97-341832/31.
DR N-PSDB; T64809.
PT Diagnosis and treatment of proliferative arterial disease - using

```

W85471;
 DT 18-FEB-1999 (first entry)
 DE ATG-750/RC-9 (allograft inflammatory factor-1).
 KW human; ATG-1120; AIF-1-delta; allograft inflammatory factor-1-delta;
 KW ATG-1; RC-9; ATG-705; chromosome 6p21.3; atherosclerosis; restenosis;
 KW hypertension; diabetes; stroke; cancer; leukaemia; autoimmune disease;
 KW psoriasis; rheumatoid arthritis; aberrant fibroproliferative;
 KW inflammatory response.
 OS Homo sapiens.
 PN EP-879883-A1.
 PD 25-NOV-1998.
 PF 23-DEC-1997; 310564.
 PR 22-MAY-1997; US-861494.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Douglas SA;
 DR WPI: 98-596875/51.
 DR N-PSDB; V82802.
 PT New nucleic acid encoding ATG-1120, a splice variant of allograft
 PT inflammatory factor - useful for treatment and diagnosis of e.g.
 PT atherosclerosis, restenosis, stroke and cancer
 PS Example 1; Page 14; 25pp; English.
 CC The present sequence represents a human ATG-705 (RC-9/AIF-1 (AIF
 CC (allograft inflammatory factor)-1-) protein. The specification
 CC describes a splice variant of ATG-1 called ATG 1120 (ATG-1-delta)
 CC which has exon 4 deleted. The ATG-1120 gene is at chromosome 6p21.3.
 CC The ATG-1120 products are used to treat ATG-1120 associated diseases
 CC such as atherosclerosis, restenosis, hypertension, diabetes, stroke,
 CC cancer (particularly leukaemia), psoriasis, autoimmune disease (e.g. human
 CC immunodeficiency virus), psoriasis, rheumatoid arthritis and other
 CC (non-)cardiovascular conditions involving aberrant fibroproliferative
 CC or inflammatory responses.
 SQ Sequence 147 AA;

Query Match 45.7%; Score 37; DB 1; Length 147;
 Best Local Similarity 50.0%; Pred. NO. 20;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LLKAEKNKXIEIN 14
 |||| : : : |||
 DB 15 LLKAOQEERLDEIN 28

RESULT 10

W85469
 ID W85469 standard; Protein; 147 AA.
 AC W85469;
 DT 18-FEB-1999 (first entry)
 DE ATG-750/RC-9 (allograft inflammatory factor-1) protein.
 KW human; ATG-1100; AIF-1gamma; allograft inflammatory factor-1gamma;
 KW ATG-1; RC-9; ATG-705; chromosome 6p21.3; atherosclerosis; restenosis;
 KW hypertension; diabetes; stroke; cancer; leukaemia; autoimmune disease;
 KW psoriasis; rheumatoid arthritis; aberrant fibroproliferative;
 KW inflammatory response.
 OS Homo sapiens.
 PN EP-879882-A1.
 PD 25-NOV-1998.
 PF 27-OCT-1997; 308557.
 PR 22-MAY-1997; US-862167.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Douglas SA;
 DR WPI: 98-596874/51.
 DR N-PSDB; V82800.
 PT New nucleic acid encoding ATG-1100, a splice variant of allograft
 PT inflammatory factor - useful for treatment and diagnosis of e.g.
 PT atherosclerosis, restenosis, stroke and cancer
 PS Example 1; Page 14; 24pp; English.
 CC The present sequence represents a human ATG-705 (RC-9/AIF-1 (AIF
 CC (allograft inflammatory factor)-1-) protein. The specification
 CC describes a splice variant of ATG-1 called ATG 1100 (ATG-1gamma)
 CC comprising exons 4, 5, and 6. The ATG-1100 gene is at chromosome 6p21.3.
 CC The ATG-1100 products are used to treat ATG-1100 associated diseases
 CC such as atherosclerosis, restenosis, hypertension, diabetes, stroke,

CC cancer (particularly leukaemia), autoimmune disease (e.g. human
 CC immunodeficiency virus), psoriasis, rheumatoid arthritis and other
 CC (non-)cardiovascular conditions involving aberrant fibroproliferative
 CC or inflammatory responses.
 SQ Sequence 147 AA;

Query Match 45.7%; Score 37; DB 1; Length 147;
 Best Local Similarity 50.0%; Pred. NO. 20;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LLKAEKNKXIEIN 14
 |||| : : : |||
 DB 15 LLKAOQEERLDEIN 28

RESULT 11

W85470
 ID W85470 standard; Protein; 133 AA.
 AC W85470;
 DT 18-FEB-1999 (first entry)
 DE ATG-1120 (allograft inflammatory factor-1-delta) protein.
 KW human; ATG-1120; AIF-1-delta; allograft inflammatory factor-1-delta;
 KW ATG-1; RC-9; ATG-705; chromosome 6p21.3; atherosclerosis; restenosis;
 KW hypertension; diabetes; stroke; cancer; leukaemia; autoimmune disease;
 KW psoriasis; rheumatoid arthritis; aberrant fibroproliferative;
 KW inflammatory response.
 OS Homo sapiens.
 PN EP-879883-A1.
 PD 25-NOV-1998.
 PF 23-DEC-1997; 310564.
 PR 22-MAY-1997; US-861494.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Douglas SA;
 DR WPI: 98-596875/51.
 DR N-PSDB; V82801.
 PT New nucleic acid encoding ATG-1120, a splice variant of allograft
 PT inflammatory factor - useful for treatment and diagnosis of e.g.
 PT atherosclerosis, restenosis, stroke and cancer
 PS Claim 1; Page 7; 25pp; English.
 CC The present sequence represents a human ATG-1120 (AIF (allograft
 CC inflammatory factor)-1-delta) protein. The ATG-1-delta cDNA
 CC sequence represents a splice variant of ATG-1 (RC-9/ATG-705) (V82801),
 CC having exon 4 deleted. The ATG-1120 gene is at chromosome 6p21.3.
 CC The products are used to treat ATG-1120 associated diseases such as
 CC atherosclerosis, restenosis, hypertension, diabetes, stroke, cancer
 CC (particularly leukaemia), autoimmune disease (e.g. human
 CC immunodeficiency virus), psoriasis, rheumatoid arthritis and other
 CC (non-)cardiovascular conditions involving aberrant fibroproliferative
 CC or inflammatory responses.
 SQ Sequence 133 AA;

Query Match 45.7%; Score 37; DB 1; Length 133;
 Best Local Similarity 50.0%; Pred. NO. 18;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LLKAEKNKXIEIN 14
 |||| : : : |||
 DB 15 LLKAOQEERLDEIN 28

RESULT 12

R43390
 ID R43390 standard; Protein; 275 AA.
 AC R43390;
 DT 08-MAY-1994 (first entry)
 DE Sequence specific to Neisseria gonorrhoeae.
 KW Discrimination; false positive.
 OS Neisseria gonorrhoeae.
 PN US5256536-A.
 PD 26-OCT-1993.
 PF 09-NOV-1990; 611528.

[illegible]

```

FT FT /label= Domain VI
FT FT /note= "predicted to form globular structure"
FT FT 55..57
FT FT /note= "N-linked glycosylation site"
FT FT 89..91
FT FT /note= "N-linked glycosylation site"
FT FT 287..527
FT FT /label= Domain V
FT FT /note= "contains four and one half Cystein-rich EGF-like
FT FT repeats, predicted to have rigid rod-like
FT FT structure"
FT FT 303..305
FT FT /note= "N-linked glycosylation site"
FT FT 363..365
FT FT /note= "N-linked glycosylation site"
FT FT 380..382
FT FT /note= "N-linked glycosylation site"
FT FT 470..472
FT FT /note= "N-linked glycosylation site"
FT FT 528..723
FT FT /label= Domain IVb
FT FT /note= "predicted to form globular structure"
FT FT 724..1175
FT FT /label= Domain IIib
FT FT /note= "contains ten and one half Cystein-rich EGF-like
FT FT repeats, predicted to have rigid rod-like
FT FT structure"
FT FT 746..748
FT FT /note= "N-linked glycosylation site"
FT FT 1061..1063
FT FT /note= "N-linked glycosylation site"
FT FT 1176..1379
FT FT /label= Domain IVa
FT FT /note= "predicted to form globular structure"
FT FT 1180..1573
FT FT /label= Domain IIIa
FT FT /note= "contains four Cystein-rich EGF-like repeats,
FT FT predicted to have rigid rod-like structure"
FT FT 1574..2153
FT FT /label= Domain I+II
FT FT /note= "forms two B-type chains, forms triple
FT FT coiled-coil structure"
FT FT 1597..1599
FT FT /note= "N-linked glycosylation site"
FT FT 1614..1616
FT FT /note= "N-linked glycosylation site"
FT FT 1700..1702
FT FT /note= "N-linked glycosylation site"
FT FT 1810..1812
FT FT /note= "N-linked glycosylation site"
FT FT 1901..1903
FT FT /note= "N-linked glycosylation site"
FT FT 1916..1918
FT FT /note= "N-linked glycosylation site"
FT FT 1920..1922
FT FT /note= "N-linked glycosylation site"
FT FT 2017..2019
FT FT /note= "N-linked glycosylation site"
FT FT 2028..2030
FT FT /note= "N-linked glycosylation site"
FT FT 2045..2047
FT FT /note= "N-linked glycosylation site"
FT FT 2126..2128
FT FT /note= "N-linked glycosylation site"
FT FT 2154..3110
FT FT /label= Domain G
FT FT /note= "forms large globule at end of laminin long arm"
FT FT 2240..2242
FT FT /note= "N-linked glycosylation site"
FT FT 2360..2362
FT FT /note= "N-linked glycosylation site"
FT FT 2435..2437
FT FT /note= "N-linked glycosylation site"

```

```

FT modified_site 2478..2480
FT /note= "N-linked glycosylation site"
FT modified_site 2551..2553
FT /note= "N-linked glycosylation site"
FT modified_site 2558..2560
FT /note= "N-linked glycosylation site"
FT modified_site 2648..2650
FT /note= "N-linked glycosylation site"
FT modified_site 2868..2870
FT /note= "N-linked glycosylation site"
FT modified_site 2893..2895
FT /note= "N-linked glycosylation site"
FT W09508628-A2.
FT 30-MAR-1995.
FT PD 21-SEP-1994; U10730.
FT PF 22-SEP-1993; US-125077.
FT PR (LJOL-) LA JOLLA CANCER RES FOUND.
FT PA Engvall E, Leivo I;
FT PI WPI; 95-139597/18.
FT DR N-PSDB; Q86480 and T17419.
FT DR New merosin fragments, corresp. DNA and antibodies - for diagnosing
FT PT tumour malignancy, promoting or inhibiting neurite growth and
FT PT promoting cell attachment.
FT PS Claim 5; Fig 6; 65pp; English.
FT CC This sequence represents the human 380-400 kD merosin major subunit.
FT CC Merosin is an isoform of laminin and shows structural and sequence
FT CC similarity to the human laminin A chain. Mature human merosin is
FT CC 30 amino acids larger than the human laminin A chain. Similarly to all
FT CC laminin chains, the merosin protein has distinct domains which are
FT CC predicted to have globular regions, cysteine-rich rod-like regions and
FT CC helical structures (see features table). Merosin has a large globular
FT CC domain at the carboxy terminal end. The merosin protein has an
FT CC apparent mol. wt. of about 800 kD and is composed of four polypeptides
FT CC with molecular weights of 300, 200, 200 and 80 kD. The 300 kD
FT CC polypeptide is joined to the 200 kD polypeptides by disulphide bonds
FT CC and the 300 and 80 kD polypeptides comprise the major subunit protein
FT CC given in R71729. Merosin is found in placenta, striated muscle,
FT CC peripheral nerve, trophoblasts and human Schwann cell neoplasms, among
FT CC other tissues. The 380-400 major merosin subunit also yields a 65 kD
FT CC subunit. The 380-400 merosin subunit has been designated merosin
FT CC polypeptide, merosin subunit, M chain or laminin M chain. Fragments
FT CC of the merosin protein may be used as antigens to raise anti-merosin
FT CC antibodies. These antibodies may be used in the detection of merosin,
FT CC as the absence of merosin in a tumour sample indicates malignancy.
FT CC Contacting a neurone with merosin promotes neurite outgrowth. The
FT CC merosin polypeptide may also be used in contacting inhibitors of neurite
FT CC outgrowth, thereby also promoting the outgrowth. Merosin also promotes
FT CC cell attachment. The merosin gene has been mapped to chromosome 6,
FT CC more precisely to bands 6q22->q23.
FT SQ Sequence 3110 AA;

```

```

Query Match 44.4%; Score 36; DB 1; Length 3110;
Best Local Similarity 42.9%; Pred. No. 8.7e+02;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 5 EXNXXIEEINTXLD 18
DB 1841 EANRLADEINSIID 1854

```

```

Search completed: November 13, 1999, 19:00:05
Job time: 144 sec

```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:56:36 ; Search time 64.87 Seconds
(without alignments)
3.347 Million cell updates/sec

Title: US-08-913-430-15
Perfect score: 81
Sequence: 1 LLKAEKXKIEINTXLDN 19

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/PCTUS9_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	92.6	627	2	US-08-703-947-2
2	39	48.1	639	1	US-08-466-390-2
3	39	48.1	639	1	US-08-470-950-2
4	39	48.1	639	2	US-08-467-781-2
5	39	48.1	639	2	US-08-195-487-2
6	39	48.1	639	2	US-08-483-924-2
7	39	48.1	639	3	PCT-US93-06160-2
8	38	46.9	2366	1	US-08-480-604A-10
9	38	46.9	2366	2	US-08-405-496A-10
10	37	45.7	421	2	US-07-955-905A-27
11	37	45.7	257	2	US-08-715-204-6
12	36	44.4	275	1	US-07-611-528A-2
13	36	44.4	275	1	US-08-083-946-2
14	36	44.4	337	1	US-08-317-223-3
15	36	44.4	3248	1	US-08-353-700-1
16	36	44.4	337	1	US-08-445-135-4
17	36	44.4	2482	1	US-08-328-254-6
18	36	44.4	3111	2	US-08-460-309-4
19	36	44.4	3111	2	US-08-125-077-4
20	36	44.4	337	3	PCT-US95-12675-3
21	36	44.4	3248	3	PCT-US95-16216-1
22	35	43.2	42	1	US-07-651-710A-23
23	35	43.2	908	2	US-08-356-354-6
24	34	42.0	33	2	US-08-669-791C-2
25	34	42.0	33	2	US-08-669-791C-5
26	34	42.0	250	2	US-08-562-311-2
27	34	42.0	37	2	US-08-669-790C-2
28	34	42.0	33	2	US-08-669-790C-3
29	34	42.0	33	2	US-08-669-790C-4
30	34	42.0	547	2	US-08-467-822-35
31	34	42.0	1045	2	US-08-553-438A-6
32	34	42.0	665	2	US-08-929-967-7
33	34	42.0	1079	2	US-08-929-967-8
34	34	42.0	33	2	US-08-808-825-5
35	34	42.0	33	2	US-08-808-825-6
36	34	42.0	33	2	US-08-808-825-7
37	34	42.0	48	2	US-08-808-825-8
38	34	42.0	573	3	PCT-US94-06362-1
39	34	42.0	573	3	PCT-US96-11373-1

Sequence 1, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 26, Appl
Sequence 34, Appl
Sequence 1, Appli

34 42.0 573 3 PCT-US96-11375-1
33 40.7 372 1 US-07-813-584A-3
32 1 US-08-330-515-3
107 2 US-08-625-209A-26
115 2 US-08-461-859-34
33 40.7 1066 2 US-08-308-818-1

ALIGNMENTS

RESULT 1
US-08-703-947-2
; Sequence 2, Application US/08703947
; Patent No. 5788962
; GENERAL INFORMATION:
; APPLICANT: Wise, Kim S.
; APPLICANT: Mcintosh, Mark A.
; TITLE OF INVENTION: DNA Sequences Coding for Mycoplasma
; TITLE OF INVENTION: Hyopneumoniae Surface Antigens,
; TITLE OF INVENTION: Corresponding Proteins and Use in
; TITLE OF INVENTION: Vaccines and Diagnostic Procedures
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace J. Fishel
; STREET: 929 Fee Fee Road, Suite 100
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63043
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
; COMPUTER: Hewlett-Packard Vectra
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,947
; FILING DATE: 28-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/373,957
; FILING DATE: January 17, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fishel, Grace J.
; REGISTRATION NUMBER: 25864
; REFERENCE/DOCKET NUMBER: UYM 8141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 878-0440
; TELEFAX: (314) 275-7693
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 amino acid residues
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: predicted amino acid sequence of complete
; DESCRIPTION: 627 residues of the P65 lipoprotein, derived
; DESCRIPTION: from the nucleic acid sequence
; HYPOTHETICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: whole polypeptide
; ORIGINAL SOURCE:
; ORGANISM: Mycoplasma hyopneumoniae
; STRAIN: J
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE: unicellular bacterium
; CELL LINE:
; ORGANELLE:

IMMEDIATE SOURCE:
LIBRARY: Genomic in Charon 4A, GEM12
CLONE: MhpJ25, MhpJ35, MhpJG35, pZJ25, pZJ25.14,
pZJG35.1, pZJG35.12, pZJG35.13, pZJG35.14
POSITION IN GENOME:
CHROMOSOME/SEGMENT: single chromosome
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: 627 amino acid sequence representing
NAME/KEY: complete sequence (including signal
NAME/KEY: sequence) of surface lipoprotein p65
LOCATION: entire derived coded sequence
IDENTIFICATION METHOD: clone identified by immunodetection of
IDENTIFICATION METHOD: protein product with antiserum specific for
IDENTIFICATION METHOD: p65; residue sequence deduced from nucleic
IDENTIFICATION METHOD: acid sequence
OTHER INFORMATION: immunogenic surface lipoprotein of no known
OTHER INFORMATION: function: C-terminus exposed on external
OTHER INFORMATION: surface of cell; N-terminal signal sequence
OTHER INFORMATION: (first 29 amino acid residues) cleaved during
OTHER INFORMATION: lipid modification process
PUBLICATION INFORMATION:
AUTHORS: Mary F. Kim, Maniagh B. Heidari, Susan J.
AUTHORS: Stull, Mark A. McIntosh, and Kim S. Wise
TITLE: Identification and Mapping of an
TITLE: Immunogenic Region of Mycoplasma
TITLE: hypneumonae p65 Surface Lipoprotein
TITLE: Expressed in Escherichia coli from a Cloned
TITLE: Genomic Fragment
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 8
PAGES: 2637-2643
DATE: August 1990
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: From 1 to 627
US-08-703-947-2

```

Query Match          92.6%  Score 75; DB 2; Length 627;
Best Local Similarity 84.2%  Pred. No. 8.7e-06;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 LLKAEXNKKXIEINTXLDN 19
Db 215 LLKAEVNKKKIEINTNLN 233

```

```

RESULT      2
US-08-466-390-2
; Sequence 2, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT:  TOUTKATLY, GARY
; APPLICANT:  LIDGARD, GRAHAM P
; TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES:  6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
; STREET:  125 HIGH STREET
; CITY:  BOSTON
; STATE:  MA
; COUNTRY:  USA
; ZIP:  02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  PatentIn Release #1.0, Version #1.25

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-390-2
;
; Query Match 48.1%; Score 39; DB 1; Length 639;
; Best Local Similarity 33.3%; Pred No. 20;
; Matches 6; Conservative 7; Mismatches 5; Indels 0
;
; QY 2 LKAEXNKXIEINTXLDN 19
; : : : : :
; DB 342 IQAEQDKRIEVRDAMEN 359
;
; RESULT 3
; US-08-470-950-2
; Sequence 2, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOURATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,950
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-950-2
;
; Query Match 48.1%; Score 39; DB 1; Length 639;
; Best Local Similarity 33.3%; Pred No. 20;

```

Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 LKAEKNXIEINTXLDN 19
: : : : :
Db 342 IQAEQDRKIEVRDAMEN 359

RESULT 4

US-08-467-781-2
; Sequence 2, Application US/08467781
; Patent No. 5780596

; GENERAL INFORMATION:
; APPLICANT: TOKUTATY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,781
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-467-781-2

Query Match 48.1%; Score 39; DB 2; Length 639;
Best Local Similarity 33.3%; Pred. No. 20;
Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 LKAEKNXIEINTXLDN 19
: : : : :
Db 342 IQAEQDRKIEVRDAMEN 359

RESULT 5

US-08-195-487-2
; Sequence 2, Application US/08195487
; Patent No. 5783403

; GENERAL INFORMATION:
; APPLICANT: TOKUTATY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA

; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R

; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-487-2

Query Match 48.1%; Score 39; DB 2; Length 639;
Best Local Similarity 33.3%; Pred. No. 20;
Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 2 LKAEKNXIEINTXLDN 19
: : : : :
Db 342 IQAEQDRKIEVRDAMEN 359

RESULT 6

US-08-483-924-2
; Sequence 2, Application US/08483924
; Patent No. 5882876

; GENERAL INFORMATION:
; APPLICANT: TOKUTATY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,924
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-924-2

; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-924-2

Query Match 48.1%; Score 39; DB 2; Length 639;
Best Local Similarity 33.3%; Pred. No. 20;
Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LKAEKXKXIEINTXLDN 19
::: : ||| : :
Db 342 IQAEQDKIEEVRDAMEN 359

RESULT 7
PCT-US93-06160-2
; Sequence 2, Application PC/TUS9306160
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06160
; FILING DATE: 19930621
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-06160-2

Query Match 48.1%; Score 39; DB 3; Length 639;
Best Local Similarity 33.3%; Pred. No. 20;
Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LKAEKXKXIEINTXLDN 19
::: : ||| : :
Db 342 IQAEQDKIEEVRDAMEN 359

RESULT 8
US-08-480-604A-10
; Sequence 10, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.

; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-604A-10

Query Match 46.9%; Score 38; DB 1; Length 2366;
Best Local Similarity 41.2%; Pred. No. 1.3e+02;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LKAEKXKXIEINTXLD 18
| | : | : | :
Db 210 LSNEYSKEIDELNTYIE 226

RESULT 9
US-08-405-496A-10
; Sequence 10, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN

```

1 / GENERAL INFORMATION:
2 /
3 / APPLICANT:
4 /
5 / TITLE OF INVENTION: RECOMBINANT 47 AND 31 kD COCOA PROTEINS AND
6 /
7 / DATE OF INVENTION: PRECURSOR
8 /
9 / NUMBER OF SEQUENCES: 28
10 /
11 / COMPUTER READABLE FORM:
12 /
13 / MEDIUM TYPE: Floppy disk
14 /
15 / COMPUTER: IBM PC compatible
16 /
17 / OPERATING SYSTEM: PC-DOS/MS-DOS
18 /
19 / SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
20 /
21 / CURRENT APPLICATION DATA:
22 /
23 / APPLICATION NUMBER: US/07/955,905A
24 /
25 / FILING DATE: 21-JAN-1993
26 /
27 / CLASSIFICATION: 435
28 /
29 / INFORMATION FOR SEQ ID NO: 27:

```

US-08-715-204-6

Query Match 45.7%; Score 37; DB 2; Length 257;
Best Local Similarity 53.3%; Pred. No. 16;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LLKAEKXNKXIEINT 15
||:|:|:|:|:|
Db 63 LLRTELKTEEST 77

RESULT 12

US-07-611-528A-2
; Sequence 2, Application US/07611528A
; Patent No. 5256536
; GENERAL INFORMATION:
; APPLICANT: Miyada, Charles G
; APPLICANT: Born, Teresa L
; TITLE OF INVENTION: Nucleotide Probe for Neisseria
; TITLE OF INVENTION: Gonorrhoea
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex Patent Law & Licensing Dept.
; STREET: 3401 Hillview Ave., P.O.Box 10850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/611.528A
; FILING DATE: 19901109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Leitereg, Theodore J
; REGISTRATION NUMBER: 28,319
; REFERENCE/DOCKET NUMBER: 27150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1091
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoea
; STRAIN: Strain #125
US-07-611-528A-2

Query Match 44.4%; Score 36; DB 1; Length 275;
Best Local Similarity 42.1%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 LLKAEKXNKXIEINTXLDN 19
||:|:|:|:|:|
Db 59 LLTANKKAIQIITDFEN 77

RESULT 13

US-08-083-946-2
; Sequence 2, Application US/08083946
; Patent No. 5525717
; GENERAL INFORMATION:
; APPLICANT: Miyada, Charles G

; APPLICANT: Born, Teresa L
; TITLE OF INVENTION: Nucleotide Probe for Neisseria
; TITLE OF INVENTION: Gonorrhoea
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex Patent Law & Licensing Dept.
; STREET: 3401 Hillview Ave., P.O.Box 10850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,946
; FILING DATE: 19930625
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/611,528
; FILING DATE: 09-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Leitereg, Theodore J
; REGISTRATION NUMBER: 28,319
; REFERENCE/DOCKET NUMBER: 27150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1091
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Neisseria gonorrhoea
; STRAIN: Strain #125
US-08-083-946-2

Query Match 44.4%; Score 36; DB 1; Length 275;
Best Local Similarity 42.1%; Pred. No. 26;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 LLKAEKXNKXIEINTXLDN 19
||:|:|:|:|:|
Db 59 LLTANKKAIQIITDFEN 77

RESULT 14

US-08-317-223-3
; Sequence 3, Application US/08317223
; Patent No. 5585267
; GENERAL INFORMATION:
; APPLICANT: Jones, Jonathan C.R.
; APPLICANT: Quaranta, Vito
; APPLICANT: Tamura, Richard
; TITLE OF INVENTION: CELLULAR ATTACHMENT TO TRANS-EPITHELIAL
; TITLE OF INVENTION: APPLIANCES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobb, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,223
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,727
; FILING DATE: 05-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/151,134
; FILING DATE: 12-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/152,460
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: DESMOS.002CPI
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; IMMEDIATE SOURCE:
; CLONE: merosin
;
US-08-317-223-3

```

```

Query Match 44.4%; Score 36; DB 1; Length 337;
Best Local Similarity 42.9%; Pred. No. 33;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 5 EXNKKXIEINTXLD 18
| | :|||: |
Db 224 EANRLADEINSIID 237

```

```

RESULT 15
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORTMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
;
US-08-353-700-1

```

```

Query Match 44.4%; Score 36; DB 1; Length 3248;
Best Local Similarity 50.0%; Pred. No. 4e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 2 LKAXNKKXIEINTXLD 17
| | | | | | | | | |
Db 1135 LAEQNKWKQKEVNDLL 1150

```

```

Search completed: November 13, 1999, 10:56:37
Job time: 1366 sec

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run On: November 13, 1999, 05:08:42 ; Search time 251.81 seconds
(without alignments)
4.777 Million cell updates/sec

Title: US-08-913-430-15

Perfect score: 81
Sequence: 1 LLKAXNKKXIEINTXLDN 19

Scoring table: BLOSUM62

Searched: 497124 seqs, 63304441 residues

Database : Pending_Patents_AA.*
1: /cgn2_6/ptodata/1/paa/PCTUS9_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	92.6	627	6	US-08-373-957-2
2	75	92.6	19	13	US-08-913-430-15
3	47	58.0	96	16	US-09-270-767-40163
4	47	58.0	96	16	US-09-270-767-55379
5	43	53.1	536	16	US-09-248-796-19182
6	43	53.1	536	19	US-60-096-409-19182
7	40	49.4	425	1	PCT-US97-14336-451
8	40	49.4	846	2	US-07-936-266A-10
9	40	49.4	2495	12	US-08-827-356-3899
10	40	49.4	425	13	US-08-911-503-451
11	40	49.4	425	13	US-08-911-503A-451
12	40	49.4	226	13	US-08-966-318-1
13	40	49.4	509	16	US-09-248-796-17566
14	40	49.4	495	16	US-09-248-796-19503
15	40	49.4	509	19	US-60-096-409-17566
16	40	49.4	495	19	US-60-096-409-19503
17	40	49.4	68	19	US-60-140-806-988
18	40	49.4	186	19	US-60-142-844-572
19	40	49.4	243	19	US-60-143-868-493

20	40	49.4	525	19	US-60-145-137-483
21	40	49.4	525	19	US-60-146-315-898
22	40	49.4	853	23	US-09-102-804-3
23	40	49.4	846	23	US-09-102-804-31
24	39	48.1	391	1	PCT-US97-14336-394
25	39	48.1	929	12	US-08-827-356-4549
26	39	48.1	391	13	US-08-911-503-394
27	39	48.1	391	13	US-08-911-503A-394
28	39	48.1	98	15	US-09-107-532-6980
29	39	48.1	1226	16	US-09-248-796-19043
30	39	48.1	919	16	US-09-248-796-19408
31	39	48.1	1260	16	US-09-248-796-20258
32	39	48.1	1226	19	US-60-096-409-19043
33	39	48.1	919	19	US-60-096-409-19408
34	39	48.1	1260	19	US-60-096-409-20258
35	39	48.1	575	23	US-09-417-507-39044
36	38	46.9	2366	1	PCT-US97-15394-10
37	38	46.9	2366	7	US-08-405-496-10
38	38	46.9	2366	7	US-08-422-711-10
39	38	46.9	2366	8	US-08-480-604-10
40	38	46.9	2366	11	US-08-704-159-10
41	38	46.9	362	12	US-08-827-356-3748
42	38	46.9	354	12	US-08-827-356-3816
43	38	46.9	2366	13	US-08-957-310-10
44	38	46.9	2366	14	US-09-084-517-10
45	38	46.9	410	19	US-60-096-409-20637

ALIGNMENTS

RESULT 1
US-08-373-957-2
; Sequence 2, Application US/08373957
; GENERAL INFORMATION: Kim S.
; APPLICANT: Wise, Kim S.
; APPLICANT: McIntosh, Mark A.
; TITLE OF INVENTION: DNA Sequences Coding for Mycoplasma
; TITLE OF INVENTION: Hypneumonae Surface Antigens,
; TITLE OF INVENTION: Corresponding Proteins and Use in Vaccines
; TITLE OF INVENTION: Corresponding Proteins and Use in Vaccines
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace J. Fishel
; STREET: 929 Fee Fee Road, Suite 100
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63043
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
; COMPUTER: Hewlett-Packard Vectra
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,957
; FILING DATE: January 17, 1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fishel, Grace J.
; REGISTRATION NUMBER: 25864
; REFERENCE/DOCKET NUMBER: UVM 8141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314) 878-0440
; TELEFAX: (314) 275-7693
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 amino acid residues
; TYPE: amino acid

Sequence 483, App
Sequence 898, App
Sequence 3, Appli
Sequence 31, Appl
Sequence 394, App
Sequence 4349, Ap
Sequence 394, App
Sequence 6980, Ap
Sequence 19043, A
Sequence 19408, A
Sequence 20258, A
Sequence 19043, A
Sequence 19408, A
Sequence 20258, A
Sequence 39044, A
Sequence 10, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 3748, Ap
Sequence 3816, Ap
Sequence 10, Appl
Sequence 10, Appl
Sequence 20637, A

STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: predicted amino acid sequence of complete
DESCRIPTION: 627 residues of the p65 lipoprotein, derived
DESCRIPTION: from the nucleic acid sequence
HYPOTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: whole polypeptide
ORIGINAL SOURCE: Mycoplasma hyopneumoniae
STRAIN: J
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE: unicellular bacterium
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY: Genomic in Charon 4A, GEM12
CLONE: MhpJ25, MhpJ35, MhpJG35, pZJ25,
CLONE: pZJ25.1, pZJ25.14, pZJG35.1, pZJG35.12,
CLONE: pZJG35.13, pZJG35.14
POSITION IN GENOME:
CHROMOSOME/SEGMENT: single chromosome
MAP POSITION: unknown
UNITS: unknown
FEATURE:
NAME/KEY: 627 amino acid sequence representing
NAME/KEY: complete sequence (including signal
NAME/KEY: sequence) of surface lipoprotein p65
LOCATION: entire derived coded sequence
IDENTIFICATION METHOD: clone identified by immunodetection of
IDENTIFICATION METHOD: protein product with antiserum specific for
IDENTIFICATION METHOD: p65; residue sequence deduced from nucleic
IDENTIFICATION METHOD: acid sequence
OTHER INFORMATION: immunogenic surface lipoprotein of no known
OTHER INFORMATION: function: C-terminus exposed on external
OTHER INFORMATION: surface of cell; N-terminal signal sequence
OTHER INFORMATION: (first 29 amino acid residues) cleaved during
OTHER INFORMATION: lipid modification process
PUBLICATION INFORMATION:
AUTHORS: Mary F. Kim, Manijeh B. Heidari, Susan J.
AUTHORS: Stull, Mark A. McIntosh, and Kim S. Wise
TITLE: Identification and Mapping of an
TITLE: Immunogenic Region of Mycoplasma
TITLE: hyopneumoniae p65 surface lipoprotein
TITLE: Expressed in Escherichia coli from a Cloned
TITLE: Genomic Fragment
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 8
PAGES: 2637-2643
DATE: August 1990
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: From 1 to 627
US-08-373-957-2

Query Match 92.6%; Score 75; DB 6; Length 627;
Best Local Similarity 84.2%; Pred. No. 0.00024;
Matches 16; Conservative: 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LKAEKNKXIEEINTXLDN 19
||||| ||| ||||| |||

Db 215 LKAEVKNKXIEEINTXLDN 233

RESULT 2

US-08-913-430-15

; Sequence 15, Application US/08913430B
; GENERAL INFORMATION:
; APPLICANT: Walker, John
; APPLICANT: Lee, Rogan
; APPLICANT: Dougherty, Stephen W.
; TITLE OF INVENTION: Antigen Composition Against Mycoplasma
; FILE REFERENCE: U-011415-0
; CURRENT APPLICATION NUMBER: US/08/913,430B
; CURRENT FILING DATE: 1997-12-09
; EARLIER APPLICATION NUMBER: PCT/AU96/00149
; EARLIER FILING DATE: 1996-03-15
; EARLIER APPLICATION NUMBER: PN 1789
; EARLIER FILING DATE: 1995-03-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 15
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Mycoplasma hyopneumoniae
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: (6)
; OTHER INFORMATION: Undetermined
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: (9)
; OTHER INFORMATION: Undetermined
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: (16)
; OTHER INFORMATION: Undetermined
; US-08-913-430-15

Query Match 92.6%; Score 75; DB 13; Length 19;
Best Local Similarity 100.0%; Pred. No. 4.2e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LKAEKNKXIEEINTXLDN 19
||||| ||||| ||||| |||

Db 1 LKAEKNKXIEEINTXLDN 19

RESULT 3

US-09-270-767-40163
; Sequence 40163, Application US/09270767
; GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40163
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: xaa means any amino acid
US-09-270-767-40163

Query Match 58.0%; Score 47; DB 16; Length 96;
Best Local Similarity 44.4%; Pred. No. 1.2;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LKAEKNKXIEEINTXLDN 19
| | | | | : : : | | |

Db 43 LMAVNSSLDLNTALDN 60

RESULT 4

US-09-270-767-55379
; Sequence 55379, Application US/09270767
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-034
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55379
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-55379

Query Match 58.0%; Score 47; DB 16; Length 96;
Best Local Similarity 44.4%; Pred. NO. 1.2;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LKAEKNKXIEEINTXLDN 19
| | | : : : : :
DB 43 LMARVNSLDDLNTALDN 60

RESULT 5
US-09-248-796-19182
; Sequence 19182, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 19182
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-19182

Query Match 53.1%; Score 43; DB 16; Length 536;
Best Local Similarity 36.8%; Pred. NO. 39;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 LKAEKNKXIEEINTXLDN 19
| | | : : : : :
DB 161 ILEAQLNKYVNDINNFNN 179

RESULT 6
US-60-096-409-19182
; Sequence 19182, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 19182
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-19182

Query Match 53.1%; Score 43; DB 19; Length 536;
Best Local Similarity 36.8%; Pred. NO. 39;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 LKAEKNKXIEEINTXLDN 19
| | | : : : : :
DB 161 ILEAQLNKYVNDINNFNN 179

RESULT 7
PCT-US97-14436-451
; Sequence 451, Application PC/TUS9714436
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,
; TITLE OF INVENTION: POLYPEPTIDES AND THEIR USES
; NUMBER OF SEQUENCES: 713
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/14436
; FILING DATE: 15-AUG-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/024,022
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmil, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50533
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 451:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 425 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US97-14436-451

Query Match 49.4%; Score 40; DB 1; Length 425;
Best Local Similarity 43.8%; Pred. NO. 94;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 AEXNKKXIEEINTXLDN 19
| | | : : : : :
DB 283 ANVNKTVEEVROHLEN 298

RESULT 8
US-07-936-266A-10
; Sequence 10, Application US/07936266A
; GENERAL INFORMATION:
; APPLICANT: Barre-Sinoussi, Francoise

```

RESULT      9
US-08-827-356-3899
; Sequence 3899, Application US/08827356
; GENERAL INFORMATION:
; APPLICANT: George H. Shimer, Jr.
; APPLICANT: George H. Miller
; APPLICANT: Roberta S. Hare
; APPLICANT: Karen J. Shaw
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS RELATED
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 574
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Schering-Plough Corporation
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07033-0530

```

Query Match 49.4%; Score 40; DB 12; Length 2495;
Best Local Similarity 57.1%; Pred. No. 7.4e+02;
Matches 8; Conservative 1; Mismatches 5; Indels

REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50533
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 451:
SEQUENCE CHARACTERISTICS:
LENGTH: 425 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-911-503A-451

Query Match 49.4%; Score 40; DB 13; Length 425;
Best Local Similarity 43.8%; Pred. No. 94;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 AEXNKXIEEINTXLDN 19
| | | | | | | | | |
Db 283 ANVKTVEEVROHLEN 298

RESULT 11
US-08-911-503A-451
; Sequence 451, Application US/08911503A
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lonetto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: NOVEL PROKARYOTIC POLYNUCLEOTIDES,
; POLYPEPTIDES AND THEIR USES
; NUMBER OF SEQUENCES: 713
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,503A
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/024,022
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50533
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 451:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 425 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-911-503A-451

Query Match 49.4%; Score 40; DB 13; Length 425;
Best Local Similarity 43.8%; Pred. No. 94;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 AEXNKXIEEINTXLDN 19
| | | | | | | | | |
Db 283 ANVKTVEEVROHLEN 298

RESULT 12
US-08-966-318-1
; Sequence 1, Application US/08966318
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN MYOSIN HEAVY CHAIN-LIKE PROTEINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/966,318
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0413 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-853-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LUNGNOT18
; CLONE: 2220246
US-08-966-318-1

Query Match 49.4%; Score 40; DB 13; Length 226;
Best Local Similarity 52.9%; Pred. No. 45;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 LKAEKXKXIEEINTXLD 18
| | | | | | | | | |
Db 180 LKAEKXKXIEEINTXLD 196

RESULT 13
US-09-248-796-17566
; Sequence 17566, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 17566
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-17566

Db 296 EFNDRIEELNEKLEN 310

Search completed: November 13, 1999, 05:08:44
Job time: 10870 sec

Query Match 49.4%; Score 40; DB 16; Length 509;
Best Local Similarity 53.3%; Pred. No. 1.2e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 EXNXXIEEINTXLDN 19
| | ||||| :|
Db 296 EFNDRIEELNEKLEN 310

RESULT 14
US-09-248-796-19503
; Sequence 19503, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 19503
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-19503

Query Match 49.4%; Score 40; DB 16; Length 495;
Best Local Similarity 42.1%; Pred. No. 1.1e+02;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 LLRAEXNKNXIEEINTXLDN 19
| | ||| :|
Db 377 LTNAEKNKQLAKLRAQLDN 395

RESULT 15
US-60-096-409-17566
; Sequence 17566, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 17566
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-096-409-17566

Query Match 49.4%; Score 40; DB 19; Length 509;
Best Local Similarity 53.3%; Pred. No. 1.2e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 EXNXXIEEINTXLDN 19
| | ||||| :|

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 12:08:09 ; Search time 75.45 Seconds
(without alignments)
10.089 Million cell updates/sec

Title: US-08-913-430-15
Perfect score: 81
Sequence: 1 LKAEYNKXIEINTXLDN 19

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	54.3	566	2 F70028	transmembrane rece
2	41.5	51.2	315	2 G70165	aldose reductase h
3	41	50.6	1095	2 P01114	SKDC25 protein -
4	40	49.4	846	1 P01JND	env polyprotein pr
5	40	49.4	390	2 JCS311	galactokinase (EC
6	40	49.4	87	2 S77861	ribosomal protein
7	40	49.4	979	2 JQ0894	PII5 protein - Myc
8	39	48.1	1938	1 A00997	myosin heavy chain
9	39	48.1	382	2 F70116	hypothetical prote
10	39	48.1	124	2 D64776	ybaJ protein - Esc
11	39	48.1	422	2 S73369	hypothetical prote
12	39	48.1	678	2 A71287	probable cytoplasm
13	38.5	47.5	674	2 G70875	probable 2 - Mycob
14	38	46.9	1173	1 VG1HHC	E2 glycoprotein pr
15	38	46.9	788	2 I40776	pflA protein - Cam
16	38	46.9	2366	2 S10317	toxin B - Clostrid
17	38	46.9	668	2 S64123	hypothetical prote
18	37.5	46.3	578	2 A64696	hypothetical prote
19	37.5	46.3	650	2 T00617	endostyle-specific
20	37	45.7	138	1 R5YM29	ribosomal protein
21	37	45.7	239	1 LKMSWA	mannose-binding le
22	37	45.7	1057	2 S72648	sucrose-phosphate
23	37	45.7	421	2 A24810	phaseolin beta cha
24	37	45.7	430	2 S10156	alpha-phaseolin pr
25	37	45.7	436	2 A23498	phaseolin type alp
26	37	45.7	294	2 A96888	lactose-binding le
27	37	45.7	576	2 A32228	flagellin A - Camp
28	37	45.7	144	2 C70437	ATP synthase F0 su
29	37	45.7	143	2 JC5246	allograft inflama
30	37	45.7	227	2 G64235	hypothetical prote
31	37	45.7	201	2 B71861	hypothetical prote
32	37	45.7	507	2 H69622	flagellar hook-ass
33	37	45.7	1217	2 T00607	probable TBP-bind
34	37	45.7	732	2 T01208	ADPglucose--starch
35	37	45.7	1697	2 T00079	hypothetical prote
36	37	45.7	433	2 S55296	sensory rhodopsin
37	36.5	45.1	369	2 G64224	cell division prot
38	36	44.4	173	1 CFYCB	C-phycocyanin beta
39	36	44.4	63	1 R5EC29	ribosomal protein

RESULT 1
F70028
transmembrane receptor taxis protein homolog yvaQ - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
C:Accession: F70028
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: F70028
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-566 <KUN>
A:Cross-references: GB:Z99121; GB:AL009126; NID:g2635827; PID:ell186057; PID:g2635882
A:Experimental source: strain 168
C:Genetics:
A:Gene: yvaQ
Query Match 54.3%; Score 44; DB 2; Length 566;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 2 LKAEYNKXIEINTXLDN 19
| | | | | | | | | |
DB 81 LQEMNQIMEDINQKLDN 98

RESULT 2
G70165
aldose reductase homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
C:Accession: G70165
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete.
A:Reference number: A70100; MUID:98065943
A:Accession: G70165
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-315 <KLE>
A:Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PID:g2688432; TIGR:BB0528
A:Experimental source: strain B31

C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 04-Sep-1998
 C:Accession: JQ0894
 R:Notarnicola, S.M.; McIntosh, M.A.; Wise, K.S.
 Gene 97, 77-85, 1991
 A:Title: A Mycoplasma hyorhinis protein with sequence similarities to nucleotide-binding
 A:Reference number: JQ0894; MUID:91138990
 A:Accession: JQ0894
 A:Molecule type: DNA
 A:Residues: 1-979 <NOT>
 A:Cross-references: GB:M34956
 A:Note: The authors translated the codon AAA for residue 956 as Leu
 C:Comment: This protein is located in the cytoplasm.
 C:Genetics:
 A:Genetic code: SGC3
 C:Superfamily: conserved hypothetical P115 protein
 C:Keywords: P-loop
 F:32-39/Region: nucleotide-binding motif A (P-loop)

Query Match 49.4%; Score 40; DB 2; Length 979;
 Best Local Similarity 47.1%; Pred. No. 85;
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 KAEKXKXIEINTXLDN 19
 ||| | : |||
 Db 277 KTEVESSIQEITSKLDN 293

RESULT 8
 A:00997
 myosin heavy chain, striated adductor muscle - scallop (Aequipecten irradians)
 N:Contains: myosin ATPase (EC 3.6.1.32)
 C:Species: Aequipecten irradians
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Feb-1998
 C:Accession: A40997; S13557
 R:Nyiray, L.; Goodwin, E.B.; Szent-Gyorgyi, A.G.
 J. Biol. Chem. 266, 18469-18476, 1991
 A:Title: Complete primary structure of a scallop striated muscle myosin heavy chain. Seq
 A:Reference number: A40997; MUID:92011595
 A:Accession: A40997
 A:Molecule type: mRNA
 A:Residues: 1-1938 <NYI>
 A:Cross-references: GB:X55714; NID:g5611; PID:g5612
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; coiled coil; hydrolase; muscle contraction; P-loop; tand
 F:86-763/Domain: myosin motor domain homology <MMOT>
 F:176-183/Region: nucleotide-binding motif A (P-loop)
 F:547-586/Region: actin binding #status predicted
 F:653-675/Region: actin binding #status predicted
 F:836-1938/Domain: coiled coil #status predicted <COI>
 F:836-1276/Region: S2
 F:1277-1938/Region: light meromyosin
 F:182/Binding site: ATP (Lys) #status predicted
 F:693,703/Active site: Cys #status predicted

Query Match 48.1%; Score 39; DB 1; Length 1938;
 Best Local Similarity 43.8%; Pred. No. 2,5e+02;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 KAEKXKXIEINTXLD 18
 :||| : |||
 Db 1133 RAELNRELELGERLD 1148

RESULT 9
 F70116
 hypothetical protein BB0134 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Dec-1998
 C:Accession: F70116
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: F70116
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-382 <KLE>
 A:Cross-references: GB:AE001125; GB:AE000783; NID:g2688021; PID:g2688034; TIGR:BB0134
 A:Experimental source: strain B31
 C:Superfamily: hypothetical protein TP0944; tetratricopeptide repeat homology
 F:10-43/Domain: tetratricopeptide repeat homology <TT1>
 F:44-77/Domain: tetratricopeptide repeat homology <TT2>

Query Match 48.1%; Score 39; DB 2; Length 382;
 Best Local Similarity 44.4%; Pred. No. 47;
 Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 LKAEKXKXIEINTXLDN 19
 |||| | : |||
 Db 188 LKAEINKNLNNTAAVDD 205

RESULT 10
 D64776
 ybaJ protein - Escherichia coli
 C:Species: Escherichia coli
 C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Jul-1998
 C:Accession: D64776
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: D64776
 A:Molecule type: DNA
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Residues: 1-124 <BLAT>
 A:Cross-references: GB:AE000152; GB:U00096; NID:g17866660; PID:g17866666; UWGP:D0461
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: ybaJ

Query Match 48.1%; Score 39; DB 2; Length 124;
 Best Local Similarity 53.3%; Pred. No. 15;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 EXNKXIEINTXLDN 19
 ||| | : |||
 Db 68 EDNKLIEQIDEYLD 82

RESULT 11
 S73369
 hypothetical protein C09_orf422 - Mycoplasma pneumoniae (ATCC 29342) (SGC3)
 C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C>Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 17-Jul-1998
 C:Accession: S73369
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
 A:Reference number: S73327; MUID:97105885
 A:Accession: S73369
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-422 <HIM>
 A:Cross-references: EMBL:AE000005; GB:U00089; NID:gl673684; PID:gl673692
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:

A:Genetic code: SGC3

Query Match 48.1%; Score 39; DB 2; Length 422;
Best Local Similarity 47.1%; Pred. No. 52;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 KAEYNKXIEEINTXLDN 19
| : ||| :
Db 313 KLKNDKIEINIFDS 329

RESULT 12
A71287
probable cytoplasmic filament protein A (cflA) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
C:Accession: A71287
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98322770
A:Accession: A71287
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-678 <COL>
A:Cross-references: GB:AE001246; GB:AE000520; NID:g3323046; PID:g3323053
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0748

Query Match 48.1%; Score 39; DB 2; Length 678;
Best Local Similarity 41.2%; Pred. No. 85;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 LLKAEYNKXIEEINTXLDN 17
| : ||| : |||
Db 41 LIEETNKILHLINTKL 57

RESULT 13
G70875
probable 2 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Nov-1998
C:Accession: G70875
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: G70875
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-674 <COL>
A:Cross-references: GB:AL010186; GB:AL123456; NID:g3261493; PID:el203948; PID:g2695959
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: fadH
C:Superfamily: NADH oxidase

Query Match 47.5%; Score 38.5; DB 2; Length 674;
Best Local Similarity 53.3%; Pred. No. 1e+02;
Matches 8; Conservative 5; Mismatches 1; Indels 1; Gaps 1;

QY 1 LLKAEYNKXIEEINTXLDN 15
| : ||| : |||

Db 323 VLKASNR-VDEINT 336

RESULT 14
VGIIHC
E2 glycoprotein precursor - human coronavirus (strain 229E)
N:Alternate names: peplomer glycoprotein; S glycoprotein; spike glycoprotein
C:Species: human coronavirus
A:Note: host Homo sapiens (man)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 05-Sep-1997
C:Accession: A34766; S05460
R:Raabe, T.; Schellie-Brinz, B.; Siddell, S.G.
J. Gen. Virol. 71, 1065-1073, 1990
A:Title: Nucleotide sequence of the gene encoding the spike glycoprotein of human cor
A:Reference number: A34766; MUID:90264837
A:Accession: A34766
A:Molecule type: mRNA
A:Residues: 1-1173 <RAA>
A:Cross-references: EMBL:X16816; NID:g58926; PID:g58927
A:Experimental source: strain 229E
R:Raabe, T.; Siddell, S.
Nucleic Acids Res. 17, 6387, 1989
A:Title: Nucleotide sequence of the human coronavirus HCV 229E mRNA 4 and mRNA 5 uni
A:Reference number: A34038; MUID:89366867
A:Accession: S05460
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1159-1173 <RA2>
A:Cross-references: EMBL:X15654; NID:g58921; PID:el3441; PID:g1334827
C:Superfamily: coronavirus E2 glycoprotein
C:Keywords: glycoprotein; transmembrane protein
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-1173/Product: E2 glycoprotein #status predicted <NAT>
F:1116-1138/Domain: transmembrane #status predicted <TMN>
F:23,62,98,147,171,176,220,243,326,333,440,464,518,538,542,568,581,587,663,671,930,10

Query Match 46.9%; Score 38; DB 1; Length 1173;
Best Local Similarity 37.5%; Pred. No. 2.1e+02;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 AEXNXXIEEINTXLDN 19
| : ||| : |||
Db 1079 AELNVTQKLTLDN 1094

RESULT 15
I40776
pflA protein - Campylobacter jejuni
C:Species: Campylobacter jejuni
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Aug-1996
C:Accession: I40776
R:Yao, R.; Burr, D.H.; Doig, P.; Trust, T.J.; Niu, H.; Guerry, P.
Mol. Microbiol. 14, 893-893, 1994
A:Title: Isolation of motile and non-motile insertional mutants of Campylobacter jeju
A:Reference number: I40776; MUID:95231295
A:Accession: I40776
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-788 <RES>
A:Cross-references: EMBL:U09019; NID:g508864; PID:g508865
C:Genetics:
A:Gene: pflA

Query Match 46.9%; Score 38; DB 2; Length 788;
Best Local Similarity 52.9%; Pred. No. 1.4e+02;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 LLKAEYNKXIEEINTXLDN 17
| : ||| : |||
Db 339 LSNEQVKNKAEIYINTIL 355

Search completed: November 13, 1999, 12:08:10
Job time: 2082 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 1999, 10:33:46 ; Search time 51.07 Seconds
(without alignments)
10.517 Million cell updates/sec

Title: US-08-913-430-15
Perfect score: 81
Sequence: 1 LKAEKXKXIEINTXLDN 19
Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	41	50.6	1095	CC25_SACKL	Q02342 saccharomyc
2	40	49.4	846	ENV_HVIND	P18799 human immun
3	40	49.4	390	GALI_STRMR	P96993 streptococc
4	40	49.4	979	P115_MYCHR	P41508 mycoplasma
5	39	48.1	678	CFPA_TREPA	O56336 treponema p
6	39	48.1	1938	MYS_AEQIR	P24733 aequipecten
7	39	48.1	384	RM02_KLULA	P48335 kluyveromyc
8	39	48.1	124	YBAJ_ECOLI	P37611 escherichia
9	38	46.9	2366	TOXB_CLODI	P18177 clostridium
10	38	46.9	1173	VGL2_CVR22	P15423 human coron
11	38	46.9	668	VGL3_YEAST	P53135 saccharomyc
12	37	45.7	147	A1F1_HUMAN	P55008 homo sapien
13	37	45.7	575	FLAA_CAMJE	P22251 campylobact
14	37	45.7	507	FLKG_BACSU	P39810 bacillus su
15	37	45.7	433	HTR2_HALVA	P42258 haloarcula
16	37	45.7	295	LEG3_CANFA	P38486 canis fami
17	37	45.7	239	MABA_MOUSE	P39039 mus musculu
18	37	45.7	436	PHSA_PRAVU	P07219 phaseolus v
19	37	45.7	421	PHSB_PRAVU	P02853 phaseolus v
20	37	45.7	138	RL29_MYCCA	P10142 mycoplasma
21	37	45.7	1057	SPS1_CITUN	O22060 citrus unsh
22	37	45.7	227	Y323_MYCGE	P47565 mycoplasma
23	37	45.7	257	Y2G1_CAEEL	P55326 caenorhabdi
24	36.5	45.1	389	FTSZ_MYCGE	P47466 mycoplasma
25	36	44.4	165	ADFI_CAEEL	Q07750 caenorhabdi
26	36	44.4	258	APAL_SALSA	P27007 salmo salar
27	36	44.4	2649	BPAL_HUMAN	Q03001 homo sapien
28	36	44.4	3210	CENF_HUMAN	P49454 homo sapien
29	36	44.4	548	HLVB_VTECH	P15492 vibrio chol
30	36	44.4	644	HS71_PICAN	P53421 picchia angu
31	36	44.4	3110	LMA2_HUMAN	P24043 homo sapien
32	36	44.4	749	MADI_YEAST	P40957 saccharomyc
33	36	44.4	1938	MYS_CAEEL	P02567 caenorhabdi
34	36	44.4	1655	N188_YEAST	P52593 saccharomyc
35	36	44.4	464	N2B_HAEIR	P46441 haematobia
36	36	44.4	170	PHCE_SYN6	P00312 synchococc
37	36	44.4	172	PHCB_SYN7	P06539 synchococc
38	36	44.4	1058	PMAL_DICDI	P54679 dictyosteli
39	36	44.4	127	RBFA_BORBU	O51742 borrella bu
40	36	44.4	2869	RBPL_PLAVB	Q00798 plasmodium
41	36	44.4	200	RHO2_SCHPO	Q10133 schizosacch
42	36	44.4	63	RL29_ACTKS	P46174 acyrthosiph
43	36	44.4	63	RL29_ECOLI	P02429 escherichia

ALIGNMENTS

```
RESULT 1
ID CC25_SACKL STANDARD; PRT; 1095 AA.
AC Q02342;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE CELL DIVISION CONTROL PROTEIN 25 (FRAGMENT).
GN CDC25.
OS SACCHAROMYCES KLUYVERI (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92354938.
RA PRIGOZY T., GONZALES E., BROEK D.:
RT Identification and analysis of a DNA fragment from Saccharomyces
RT kluyveri that can complement the loss of CDC25 function in
RT Saccharomyces cerevisiae.
RL GENE 117:67-72(1992).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,
CC THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.
CC -!- SIMILARITY: TO OTHER GUANINE-NUCLEOTIDE RELEASING FACTORS OF THE
CC CDC25 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M02964; G171187;
CC PIR; PC1114; PC1114.
CC PROSITE; PS00720; GDS.CDC25; 1.
CC PFAM; PF00617; RasGEF; 1.
CC PFAM; PF00618; RasGEF; 1.
CC GUANINE-NUCLEOTIDE RELEASING FACTOR; CELL DIVISION; CELL CYCLE;
CC MITOSIS; TRANSMEMBRANE.
CC NON_TER 1
CC TRANSMEM 959 980 POTENTIAL.
CC SEQUENCE 1095 AA; 125605 MW; 7665862 CRC32;

Query Match 50.6%; Score 41; DB 1; Length 1095;
Best Local Similarity 46.7%; Pred. No. 35;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 EXNKKXIEINTXLDN 19
| :|||:|:|
Db 929 EYKKLLEELTMDS 943

RESULT 2
ENV_HVIND STANDARD; PRT; 846 AA.
AC P18799;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: GP120; GP41].
GN ENV.
OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (NDK ISOLATE) (HIV-1).
OC VIRUSES; RETROVIRIDAE; RETROVIRUS.
```

RN SEQUENCE FROM N.A.
RX MEDLINE; 90034200.
RA SPIRE B., SIRE J., ZACHAR V., REY F., BARRE-SINOSSI F., GALIBERT F.,
HAMPE A., CHERMANN J.C.,
RT "Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the
human immunodeficiency virus."
RL GENE 81:275-284(1989).
CC -1- NDK, ISOLATED FROM A ZAIRIAN PATIENT AFFECTED WITH AIDS, AND IS
CC A HIGHLY CYTOPATHOGENIC STRAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M27323; G328162; -
DR PIR: JQ0066; VCLJND.
DR HIV: M27323; ENV5NDK.
DR PFAM: PF00516; GP120; 1.
DR PFAM: PF00517; GP41; 1.
KW AIDS; COAT PROTEIN; POLYPROTEIN; GLYCOPROTEIN; TRANSMEMBRANE;
KW SIGNAL.
FT SIGNAL
FT CHAIN 1 29
FT CHAIN 30 501
FT CHAIN 502 846
FT DISULFID 53 73
FT DISULFID 118 200
FT DISULFID 125 191
FT DISULFID 130 152
FT DISULFID 213 242
FT DISULFID 223 234
FT DISULFID 291 328
FT DISULFID 374 435
FT DISULFID 381 408
FT CARBOHYD 87 87
FT CARBOHYD 129 129
FT CARBOHYD 151 151
FT CARBOHYD 179 179
FT CARBOHYD 182 182
FT CARBOHYD 229 229
FT CARBOHYD 236 236
FT CARBOHYD 257 257
FT CARBOHYD 271 271
FT CARBOHYD 284 284
FT CARBOHYD 290 290
FT CARBOHYD 351 351
FT CARBOHYD 382 382
FT CARBOHYD 388 388
FT CARBOHYD 392 392
FT CARBOHYD 395 395
FT CARBOHYD 401 401
FT CARBOHYD 438 438
FT CARBOHYD 451 451
FT CARBOHYD 452 452
FT CARBOHYD 501 601
FT CARBOHYD 606 606
FT CARBOHYD 615 615
FT CARBOHYD 627 627
SQ SEQUENCE 846 AA; 96476 MW; 594ABBA1 CRC32;

Query Match 49.4%; Score 40; DB 1; Length 846;
Best Local Similarity 41.2%; Pred. No. 39;
Matches 7; Conservative 5; Mismatches 5; Indels 5; Gaps 0;

QY 3 KAEKXKXIEEINTXLDN 19

Db 332 RAENKALQVATKLN 348

RESULT 3
GALLI_STEMU
ID GALLI_STEMU STANDARD; PRT; 390 AA.
AC P96993;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GALACTOKINASE (EC 2.7.1.6).
GN GALK.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INGBRITT;
RX MEDLINE; 97128818.
RA AJDJIC D., SUTCLIFFE I.C., RUSSELL R.R.B., FERRETTI J.J.;
RT "Organization and nucleotide sequence of the Streptococcus mutans
galactose operon."
RL GENE 180:137-144(1996).
CC -1- CATALYTIC ACTIVITY: ATP + D-GALACTOSE -> ADP + D-GALACTOSE
CC -1- PHOSPHATE.
CC -1- PATHWAY: FIRST REACTION OF GALACTOSE METABOLISM (LELOIR PATHWAY).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC AND EUKARYOTIC GALACTOKINASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U21942; G1877422; -
DR PROSITE; PS00106; GALACTOKINASE; 1.
DR PROSITE; PS00627; GHMP_KINASES; 1.
DR PFAM: PF00288; GHMP_Kinases; 1.
KW TRANSFERASE; KINASE; GALACTOSE METABOLISM; ATP-BINDING.
FT NP_BIND 122 132 ATP (POTENTIAL).
SQ SEQUENCE 390 AA; 43324 MW; 304E12B2 CRC32;
Query Match 49.4%; Score 40; DB 1; Length 390;
Best Local Similarity 43.8%; Pred. No. 18;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 3 KAEKXKXIEEINTXLD 18
Db 228 RTECEKAVEELNVLLD 243
RESULT 4
P115_MYCHR
ID P115_MYCHR STANDARD; PRT; 979 AA.
AC P41508;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE P115 PROTEIN.
OS MYCOPLASMA HYORHINIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLICUTES;
OC MYCOPLASMACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 91138990.
RA TARNICOLA S.M., MCINTOSH M.A., WISE K.S.;
RT "A Mycoplasma hyorhinis protein with sequence similarities to
nucleotide-binding enzymes."
RL GENE 97:77-85(1991).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X55714; G5612; -
PIR: S13557; S13557.
PIR: A40997; A40997.
PDB: 1SCM; 30-APR-94.
PDB: 1WDC; 11-JUL-96.
PDB: PF00063; myosin_head; 1.
PFAM: PF00612; IQ; 1.
MYOSIN: MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;
ATP-BINDING; ALKYLATION; HEPTAD REPEAT PATTERN: 3D-STRUCTURE.
DOMAIN 1 835
DOMAIN 836 1938
DOMAIN 836 1938
DOMAIN 176 183
ATP (BY SIMILARITY).
MOD_RES 693 703
ALKYLATION (SH-1) (BY SIMILARITY).
MOD_RES 703 703
ALKYLATION (SH-2) (BY SIMILARITY).
HELI 778 821
TURN 822 823
TURN 825 833
TURN 834 835
SEQUENCE 1938 AA; 222821 MW; 98F6787F CRC32;

Query Match 48.1%; Score 39; DB 1; Length 1938;
Best Local Similarity 43.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 3 KAEXNKKXIEINTXLD 18
||| : ||| : ||
DB 1133 RAELENLEELGERLD 1148

RESULT 7
ID RM02_KLULA STANDARD; PRT; 364 AA.
AC P48535;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L2 PRECURSOR.
GN MRPL2 OR MRP7.
OS KLUYVEROMYCES LACTIS (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; KLUYVEROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA PAN C., SIRUM-CONNOLLY K., MASON T.L.;
RT "Essential features of the peptidyl transferase center in the yeast
mitochondrial ribosome."
RL (IN) NIERHAUS K.H. (EDS.);
RL THE TRANSLATIONAL APPARATUS, PP. 587-598, PLENUM PRESS,
NEW YORK (1993).
CC -!- FUNCTION: COMPONENT OF THE LARGE SUBUNIT OF MITOCHONDRIAL
RIBOSOME.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -!- SIMILARITY: BELONGS TO THE L27P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: U38369; G1022902; -
PROSITE: PS00831; RIBOSOMAL_L27; 1.
DR PFAM: PF01016; RIBOSOMAL_L27; 1.
KW RIBOSOMAL PROTEIN; MITOCHONDRION; TRANSIT PEPTIDE.

FT TRANSIT 1 19 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 20 364 60S RIBOSOMAL PROTEIN L2.
SQ SEQUENCE 364 AA; 42720 MW; 783232A9 CRC32;

Query Match 48.1%; Score 39; DB 1; Length 364;
Best Local Similarity 47.1%; Pred. No. 24;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 2 LKAEXNKKXIEINTXLD 18
||| : ||| : ||
DB 164 LKAESKIIISDLKVELD 180

RESULT 8
YBAJ_ECOLI STANDARD; PRT; 124 AA.
ID YBAJ_ECOLI
AC P37611; P75708;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 14.6 KD PROTEIN IN HHA-ACRB INTERGENIC REGION.
GN YBAJ.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEM G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GORDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL SCIENCE 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 19-124 FROM N.A.
RC STRAIN-K12 / C600;
RX MEDLINE; 92065825.
RA NIETO J.M., CARMONA M., BOLLAND S., JUBETE Y., DE LA CRUZ F.,
RA JUAREZ A.;
RT "The hha gene modulates haemolysin expression in Escherichia coli."
RL MOL. MICROBIOL. 5:1285-1293(1991).
RN [3]
RP IDENTIFICATION.
RX MEDLINE; 95075659.
RA BORODOVSKY M., RUDD K.E., KOONIN E.V.;
RT "Intrinsic and extrinsic approaches for detecting genes in a
bacterial genome."
RL NUCLEIC ACIDS RES. 22:4756-4767(1994).
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
FRAMESHIFTS IN POSITION 43 AND 53.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: AF000152; G1786666; -
EMBL: X57977; -; NOT_ANNOTATED_CDS.
DR ECOGENE; EGI2429; YBAJ.
KW HYPOTHETICAL PROTEIN.

SQ SEQUENCE 124 AA; 14557 MW; B8F5907F CRC32;

Query Match 48.1%; Score 39; DB 1; Length 124;
Best Local Similarity 53.3%; Pred. No. 8;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```
QY 5 EXNKKXIEINTXLDN 19
DB 68 EDNKLIEIDEYLD 82

RESULT 9
TOXB_CLODI STANDARD; PRT: 2366 AA.
AC P18177;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TOXIN B.
GN CLOSTRIDIUM DIFFICILE.
OS CLOSTRIDIUM DIFFICILE.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA MEDLINE; 90326540.
RA BARROSO L.A., WANG S.Z., PHELPS C.J., JOHNSON J.L., WILKINS T.D.;
RT "Nucleotide sequence of Clostridium difficile toxin B gene.";
RL NUCLEIC ACIDS RES. 18:4004-4004(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA VON EICHEL-STREIBER C.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 1271-2366 FROM N.A.
RC STRAIN=VPI 10463;
RA EICHEL-STREIBER C., LAUFENBERG-FELDMANN R., SARTINGEN S.,
RA SCHULZE J., SAUERBORN M.;
RT "Comparative sequence analysis of the Clostridium difficile toxins A
and B.";
RL MOL. GEN. GENET. 233:260-268(1992).
CC -!- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
CC ENTEROTOXIN CALLED A AND CYTOTOXIN B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53138; G40443; -
DR EMBL; X92982; E212009; -
DR EMBL; X60984; G40446; -
DR PIR; S10317; S10317.
KW CYTOTOXIN; TOXIN.
SQ SEQUENCE 2366 AA; 269709 MW; 39B3C774 CRC32;

Query Match 46.9%; Score 38; DB 1; Length 2366;
Best Local Similarity 41.2%; Pred. No. 2.4e+02;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 LKAEKNXIEINTXLD 18
DB 210 LSNEYSKEIDELNTYIE 226

RESULT 10
VGL2_CVH22 STANDARD; PRT: 1173 AA.
AC P15423;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
```

```
DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN).
GN S.
OS HUMAN CORONAVIRUS (STRAIN 229E).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; NIDOVIRALES;
OC CORONAVIRIDAE; CORONAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90264837.
RA RAABE T., SCHELLE-PRINZ B., SIDDELL S.G.;
RT "Nucleotide sequence of the gene encoding the spike glycoprotein of
human coronavirus HCV 229E.";
RL J. GEN. VIROL. 71:1065-1073(1990).
CC -!- FUNCTION: THE PELOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC AND IN SYNCTIUM FORMATION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16816; G58927; -
DR PIR; A34766; VGIIHC.
KW GLYCOPROTEIN; ENVELOPE PROTEIN; TRANSMEMBRANE; SIGNAL.
FT SIGNAL 1 15
FT CHAIN 16 1173 SPIKE E2 GLYCOPROTEIN.
FT DOMAIN 16 1115 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1116 1135 POTENTIAL.
FT DOMAIN 1136 1173 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1136 1157 CYS-RICH.
FT CARBOHYD 23 23 POTENTIAL.
FT CARBOHYD 62 62 POTENTIAL.
FT CARBOHYD 98 98 POTENTIAL.
FT CARBOHYD 147 147 POTENTIAL.
FT CARBOHYD 171 171 POTENTIAL.
FT CARBOHYD 176 176 POTENTIAL.
FT CARBOHYD 220 220 POTENTIAL.
FT CARBOHYD 243 243 POTENTIAL.
FT CARBOHYD 326 326 POTENTIAL.
FT CARBOHYD 333 333 POTENTIAL.
FT CARBOHYD 440 440 POTENTIAL.
FT CARBOHYD 464 464 POTENTIAL.
FT CARBOHYD 518 518 POTENTIAL.
FT CARBOHYD 538 538 POTENTIAL.
FT CARBOHYD 542 542 POTENTIAL.
FT CARBOHYD 568 568 POTENTIAL.
FT CARBOHYD 581 581 POTENTIAL.
FT CARBOHYD 587 587 POTENTIAL.
FT CARBOHYD 663 663 POTENTIAL.
FT CARBOHYD 671 671 POTENTIAL.
FT CARBOHYD 930 930 POTENTIAL.
FT CARBOHYD 1015 1015 POTENTIAL.
FT CARBOHYD 1020 1020 POTENTIAL.
FT CARBOHYD 1037 1037 POTENTIAL.
FT CARBOHYD 1049 1049 POTENTIAL.
FT CARBOHYD 1061 1061 POTENTIAL.
FT CARBOHYD 1066 1066 POTENTIAL.
FT CARBOHYD 1076 1076 POTENTIAL.
FT CARBOHYD 1082 1082 POTENTIAL.
FT CARBOHYD 1096 1096 POTENTIAL.
FT CARBOHYD 1173 1173 AA; 128639 MW; 53961EC2 CRC32;
SQ SEQUENCE
```

```
Query Match 46.9%; Score 38; DB 1; Length 1173;
Best Local Similarity 37.5%; Pred. No. 1.2e+02;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 AEXNKKXIEINTXLDN 19
DB 1173 AA; 128639 MW; 53961EC2 CRC32;
```



```

Db 1079 AELNTYVQKQLTLDN 1094
RESULT 11
YGL3_YEAST
ID YGL3_YEAST STANDARD; PRT; 668 AA.
AC P53135;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 77.3 KD PROTEIN IN SNF4-TAF60 INTERGENIC REGION.
GN YGL113W OR G3980.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA LAQUIN G.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE OF 492-668 FROM N.A.
RX MEDLINE; 97197974.
RA PAOLUZI S., MINENKOVA O., CASTAGNOLI L.;
RT "The genes encoding the transcription factor YTAII60, the G4p1
RT protein and a putative glucose transporter are contained in a 12.3 kb
RT DNA fragment on the left arm of Saccharomyces cerevisiae chromosome
RT VII."
RL YEAST 13:85-91(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z72635; E243341; -
DR EMBL; Z72634; E312762; -
DR EMBL; X97644; E240083; -
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 276 279 POLY-LEU.
SQ SEQUENCE 668 AA; 77297 MW; 3B795D20 CRC32;
Query Match 46.98; Score 38; DB 1; Length 668;
Best Local Similarity 53.38; Pred. No. 65;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Qy 4 AEXNKXIEEINTXLD 18
:|:|:|:|:|
Db 324 SEKDGIERTITSLD 338
RESULT 12
AIF1_HUMAN
ID AIF1_HUMAN STANDARD; PRT; 147 AA.
AC P35008;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ALLOGRAFT INFLAMMATORY FACTOR-1 (AIF-1) (IONIZED CALCIUM BINDING
DE ADAPTER MOLECULE 1).
GN AIF1 OR IBAL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA UTANS U., QUIST W.C., MCMAHON B.M., WILSON J.E., ARCECI R.J.,
RA RUSSELL M.E.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97069649.
RT "cDNA cloning of human allograft inflammatory factor-1: tissue
RT distribution, cytokine induction, and mRNA expression in injured rat
RT carotid arteries."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 228:29-37(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA IMAI Y., OHSAWA K., KOHSAKA S.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: MAY PLAY A ROLE IN MACROPHAGE ACTIVATION AND FUNCTION.
CC -!- PTM: PHOSPHORYLATED ON SERINE.
CC -!- SIMILARITY: SOME, TO EF-HAND CALCIUM BINDING PROTEINS.
CC -!- SIMILARITY: STRONG, TO BALLOON ANGIOPLASTY RESPONSIVE TRANSCRIPT 1
CC (BARI-1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19713; G1122909; -
DR EMBL; U49392; G1229022; -
DR EMBL; D86438; G1596163; -
MIM; 601833; -
DR HSSP; P02593; ICDM.
KW CALCIUM-BINDING; PHOSPHORYLATION.
FT CA_BIND 58 69 POTENTIAL.
FT DOMAIN 94 105 ANCESTRAL CALCIUM SITE.
FT CONFLICT 33 33 D->H (IN REF. 2).
FT CONFLICT 131 147 KEKPTGPPAKKAISELP -> RKTNTPPSQESPI (IN
FT REF. 2).
SQ SEQUENCE 147 AA; 16703 MW; BF565FEF CRC32;
Query Match 45.78; Score 37; DB 1; Length 147;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Qy 1 LKAEYNKXIEEIN 14
||||:|:|:|
Db 15 LKAEQEEERLDEIN 28
RESULT 13
FLAA_CAMJE
ID FLAA_CAMJE STANDARD; PRT; 575 AA.
AC P22251;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE FLAGELLIN A.
GN FLAA.
OS CAMPYLOBACTER JEJUNI.
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;
OC CAMPYLOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-81116;
RX MEDLINE; 91009243.
RA NUIJTEN P.J., VAN ASTEN F.J., GAASTRA W., VAN DER ZEIJST B.A.;
RT "Structural and functional analysis of two Campylobacter jejuni
RT flagellin genes."
RL J. BIOL. CHEM. 265:17798-17804(1990).
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -!- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.
CC -!- SIMILARITY: TO OTHER BACTERIAL FLAGELLINS.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; J05635; G144198; -
 CC PIR; A39228; A39228.

DR PFAM; PF00669; Flagellin_N; 1.
 DR PFAM; PF00700; Flagellin_C; 1.

KW FLAGELLA.

FT INITMET 0 0 BY SIMILARITY.

SQ SEQUENCE 575 AA; 59398 MW; 9735ED18 CRC32;

Query Match 45.7%; Score 37; DB 1; Length 575;

Best Local Similarity 47.4%; Pred. No. 81;

Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

QY 1 LLKAEKNKXIEEINTXLDN 19

Db 110 MIQADINKLMEE---LDN 124

RESULT 14

FLGK_BACSU

ID FLGK_BACSU STANDARD; PRT; 507 AA.

AC P39810;

DT 01-FEB-1995 (REL. 31, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE FLAGELLAR HOOK-ASSOCIATED PROTEIN 1 (HAP1).

GN FLGK.

OS BACILLUS SUBTILIS.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

CC BACILLUS.

CC [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA SOLDO B., LAZAREVIC V., MAUEL C., KARAMATA D.;

RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

CC [2]

RP SEQUENCE OF 1-129 FROM N.A.

RC STRAIN=168;

RX MEDLINE: 94321318.

RA MIREL D.B., LAUER P., CHAMBERLIN M.J.;

RT "Identification of flagellar synthesis regulatory and structural

RT genes in a sigma D-dependent operon of *Bacillus subtilis*."

RL J. BACTERIOL. 176:4492-4500(1994).

CC -1- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS

CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U56901; G1762338; -

DR EMBL; L14437; G451871; -

DR EMBL; 299122; E118447; -

DR SUBTILIST; BG10401; FLGK.

DR PROSITE; PS00588; FLAGELLA_BB_ROD; 1.

DR PFAM; PF00460; flg_bb_rod; 1.

KW FLAGELLA.

SQ SEQUENCE 507 AA; 54355 MW; 5C9952C8 CRC32;

Query Match

45.7%; Score 37; DB 1; Length 507;

Best Local Similarity 43.8%; Pred. No. 71;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 LKAEKNKXIEEINTXL 17

Db 169 LKAEKNTTVDVNSLL 184

RESULT 15

HTR2_HALVA

ID HTR2_HALVA STANDARD; PRT; 433 AA.

AC P42258;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE SENSORY RHODOPSIN II TRANSDUCER (HTR-II) (METHYL-ACCEPTING PHOTOTAXIS

DE PROTEIN II) (MPP-II) (FRAGMENT).

GN HTRII.

OS HALOARCUA VALLISMORTIS.

OC ARCHAEA; EURYARCHAEOTA; HALOBACTERIALES; HALOBACTERIACEAE; HALOARCUA.

CC [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29715;

RX MEDLINE: 95224074.

RA SEIDEL R., SCHARF B., GAUTEL M., KLEINE K., OESTERHELT D.;

RA ENGELHARD M.;

RT "The primary structure of sensory rhodopsin II: a member of an

RT additional retinal protein subgroup is coexpressed with its

RT transducer, the halobacterial transducer of rhodopsin II."

RL PROC. NATL. ACAD. SCI. U.S.A. 92:3036-3040(1995).

CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR

CC SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO

CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.

CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; 235308; G516322; -

DR PFAM; PF00015; MCPsignal; 1.

KW TRANSDUCER; PHOTORECEPTOR; TRANSMEMBRANE; METHYLATION.

FT NON_TER 1 1

SQ SEQUENCE 433 AA; 45935 MW; E4ABF83E CRC32;

Query Match

45.7%; Score 37; DB 1; Length 433;

Best Local Similarity 41.2%; Pred. No. 60;

Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LKAEKNKXIEEINTXLD 18

Db 211 IEAETGETVEINALDD 227

Search completed: November 13, 1999, 10:33:48
 Job time: 5198 sec

Result No.	Query		Score	Match	Length	DB	ID	Description
	No.	%						
1	75	92.6	627	2	O30643	O30643	mycoplasma	
2	75	92.6	625	2	O30704	O30704	mycoplasma	
3	45	55.6	2003	5	O22869	O22869	caenorhabdi	
4	45	55.6	2020	5	Q19658	Q19658	caenorhabdi	
5	45	55.6	2971	8	Q32065	Q32065	chlamydomon	
6	44	54.3	566	2	Q32239	Q32239	bacillus stu	
7	43	53.1	224	2	P72360	P72360	staphylococ	
8	43	53.1	785	3	O94070	O94070	candida alb	
9	43	53.1	3724	5	O77320	O77320	plasmodium	
10	42	51.9	639	9	O21970	O21970	bacterioph	
11	41.5	51.2	315	2	O51478	O51478	borrelia bu	
12	41	50.6	198	2	P70812	P70812	borrelia af	
13	41	50.6	227	3	O13950	O13950	schizosacch	
14	40	49.4	87	2	O48956	O48956	mycoplasma	
15	40	49.4	216	2	O32610	O32610	borrelia tu	
16	40	49.4	1106	5	O63310	O63310	caenorhabdi	
17	39	48.1	422	2	P75451	P75451	mycoplasma	
18	39	48.1	382	2	O51159	O51159	borrelia bu	
19	39	48.1	758	4	O15092	O15092	borrelia bu	
20	39	48.1	687	4	O75299	O75299	homo sapien	
21	39	48.1	758	4	O16891	O16891	homo sapien	
22	39	48.1	817	5	O94232	O94232	caenorhabdi	
23	39	48.1	712	5	Q20323	Q20323	caenorhabdi	
24	39	48.1	1941	5	O26079	O26079	placopecten	
25	39	48.1	1950	5	O26080	O26080	placopecten	
26	39	48.1	1951	5	O17042	O17042	acquipecten	
27	39	48.1	3394	5	O77384	O77384	plasmodium	
28	38.5	47.5	674	2	O50431	O50431	mycobacteri	
29	38	46.9	788	2	Q46092	Q46092	campylobact	

RESULT	2
O30704	
ID	O30704 PRELIMINARY; PRT; 625 AA.
AC	O30704;
DT	01-JAN-1998 (TREMELrel. 05, Created)
DT	01-JAN-1998 (TREMELrel. 05, Last sequence update)
DT	01-MAY-1999 (TREMELrel. 10, Last annotation update)
DE	SURFACE LIPOPROTEIN PRECURSOR.
OS	Mycoplasma hyopneumoniae.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes; Mycoplastmataceae; Mycoplasma.
RN	[1]
RC	SEQUENCE FROM N.A.
RC	STRAIN=232A;
RU	HSU T.; MINION F.C.;
RL	Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AF015665; AAB70214.1; -.
RF	PFW; PF00657; Lipase_GDSL; 1.
KW	Signal; Lipoprotein.

FT SIGNAL 1 31 POTENTIAL
FT CHAIN 32 625 MATURE SURFACE LIPOPROTEIN.
SQ SEQUENCE 625 AA; 70914 MW; C109763D CRC32;

Query Match 92.6%; Score 75; DB 2; Length 625;
Best Local Similarity 84.2%; Pred. No. 0.001;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 LKAEKXKXIEINTXLDN 19
||||| ||| ||||| |||
Db 215 LKAEVKNKXIEINTNLDN 233

RESULT 3
Q22869 PRELIMINARY; PRT; 2003 AA.
AC Q22869;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE NON-MUSCLE MYOSIN HEAVY CHAIN II.
GN NMV-2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2 BRISTOL;
RX MEDLINE; 96310865.
RA GUO S., KEMPHUES K.J.;
RT "A non-muscle myosin required for embryonic polarity in
Caenorhabditis elegans";
RL Nature 382:453-456 (1996).
DR EMBL; U49263; AAC47238.1; -;
DR PFAM; PF00063; myosin_head; 1.
SQ SEQUENCE 2003 AA; 231304 MW; 5D2513DD CRC32;

Query Match 55.6%; Score 45; DB 5; Length 2003;
Best Local Similarity 47.1%; Pred. No. 82;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 LKAEKXKXIEINTXLD 18
:|:|:| | |||:|:|
Db 1207 MKAQNKXIEINLTD 1223

RESULT 4
Q19658 PRELIMINARY; PRT; 2020 AA.
AC Q19658; Q22673;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE F20G4.3 PROTEIN.
GN F20G4.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWKNEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL Nature 368:32-38 (1994).
DR EMBL; 275550; CAA99931.1; -;
DR EMBL; 275538; CAA99931.1; JOINED.
DR EMBL; 275538; CAA99841.1; -;
DR EMBL; 275550; CAA99841.1; JOINED.
DR PFAM; PF00063; myosin_head; 1.
SQ SEQUENCE 2020 AA; 233418 MW; F3657A6A CRC32;

Query Match 55.6%; Score 45; DB 5; Length 2020;
Best Local Similarity 47.1%; Pred. No. 83;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 LKAEKXKXIEINTXLD 18
:|:|:| | |||:|:|
Db 1224 MKAQNKXIEINLTD 1240

RESULT 5
Q32065 PRELIMINARY; PRT; 2971 AA.
AC Q32065;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 07, Last annotation update)
DE HYPOTHETICAL 341.7 KD PROTEIN.
OS Chlamydomonas reinhardtii.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=137C;
RA WATSON A.T., PURTON S.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U62943; AAB05800.1; -;
KW Hypothetical protein; Chloroplast.
SQ SEQUENCE 2971 AA; 341665 MW; 5AC1ED04 CRC32;

Query Match 55.6%; Score 45; DB 8; Length 2971;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 LKAEKXKXIEINTXLDN 19
:|:|:| | |||:|:|
Db 1800 IKAEQNSLIENISNDFN 1817

RESULT 6
Q32239 PRELIMINARY; PRT; 566 AA.
AC Q32239;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE YVAQ PROTEIN.
GN YVAQ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE; 98044033.
RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,

BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
 BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 CHOI S.K., CODANI J.J., CONNERTON I.F., CUMINGS N.J., DANIEL R.A.,
 DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRICH S.D., EMMERSON P.T.,
 ENIAT C.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
 FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
 GHIM S.Y., GLASER P., GOFEAU A., GOLIGHTLY E.J., GRANDI G.,
 GUISSEPI G., GUY B.J., HAGA K., HALECH J., HARWOOD C.R., HENAUT A.,
 HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
 JORIS B., KARAWATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
 KOBAYASHI Y., KOETTER P., KINGSTON G., KROGH S., KUMANO M.,
 KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 MEDINA N., MELIADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
 PARRO V., POHL T.M., PORTELELLA D., POWOLLIK S., PRESCOTT A.M.,
 PRESCHAN E., PUJIC P., PURNELLE B., RAPPOPORT G., REY M., REYNOLDS S.,
 RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
 SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
 SEKIGUCHI J., SEKOWSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B.,
 SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTII A.,
 VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZNEGER T.,
 WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
 "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis",
 Nature 390:249-256(1997).
 RL Nature 390:249-256(1997).
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 299121; CAB15374.1; -;
 DR PFAM: PF00672; DUF5; 1.
 DR PFAM: PF00015; MCPsignal; 1.
 SQ SEQUENCE 566 AA; 62830 MW; 064C7A94 CRC32;

Query Match 54.3%; Score 44; DB 2; Length 566;
 Best Local Similarity 50.0%; Pred. No. 37;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LKAEKXKXIEINTXLDN 19
 I : : : : :
 Db 81 LDQENQIMEDINQKLDN 98

RESULT 7
 ID P72360 PRELIMINARY; PRT; 224 AA.
 AC P72360;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE SCDA.
 GN SCDA.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NTC 8325-4;
 RX MEDLINE: 97453472.
 RA BRUNSKILL E.W., DE JONGE B.L.M., BAYLES K.W.;
 RT "The Staphylococcus aureus scdA gene: a novel locus that affects cell
 division and morphogenesis";
 RL Microbiology 143:0-0(0).
 DR EMBL: U57060; AAB81287.1; -;
 SQ SEQUENCE 224 AA; 25398 MW; 957442BE CRC32;

Query Match 53.1%; Score 43; DB 2; Length 224;
 Best Local Similarity 36.8%; Pred. No. 22;
 Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 1 LLKAEKXKXIEINTXLDN 19
 I : : : : :
 Db 144 LIKYEQGEVDDINTVIDD 162

RESULT 8
 ID O94070 PRELIMINARY; PRT; 785 AA.
 AC O94070;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN CA9C4.16.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Candidaceae; Candida.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RA MURPHY L., HARRIS D.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RA BARRELL B.G., RAJANDREAM M.A.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RA TAIT E., SIMON M.C., KING S., BROWN A.J., GOW N.A.R., SHAW D.J.;
 RT "A Candida albicans Genome Project: Cosmid Contigs, Physical Mapping,
 and Gene Isolation";
 RL Fungal Genet. Biol. 21:308-314(1997).
 DR EMBL: AL033503; CA22026.1; -;
 SQ SEQUENCE 785 AA; 92197 MW; 623F2F22 CRC32;

Query Match 53.1%; Score 43; DB 3; Length 785;
 Best Local Similarity 36.8%; Pred. No. 70;
 Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 LLKAEKXKXIEINTXLDN 19
 I : : : : :
 Db 224 ILEAQLNKYVDINNIFNN 242

RESULT 9
 ID O77320 PRELIMINARY; PRT; 3724 AA.
 AC O77320;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE MAL3P3.3 PROTEIN.
 GN MAL3P3.3.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MUNGALL K., LAWSON D., BARRELL B.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z98547; CAB1104.1; -;
 SQ SEQUENCE 3724 AA; 448199 MW; 837A7BA3 CRC32;

Query Match 53.1%; Score 43; DB 5; Length 3724;
 Best Local Similarity 47.4%; Pred. No. 2.9e+02;
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Query Match 51.2%; Score 41.5; DB 2; Length 315;
 Best Local Similarity 61.1%; Pred. No. 51;
 Matches 11; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 LKAEKXKXIEINTXLDN 19
 DB 753 LVKADKNEISENINKICDN 771

RESULT 10
 ID 021970 PRELIMINARY; PRT; 639 AA.
 AC 021970;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
 DE DARA PROTEIN.
 GN DARA.
 OS Bacteriophage P1.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA IIDA S., HESTAND-NAUER R., SANDMEIER H., LEHNHERR H., ARBER W.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 84057693.
 RA HESTAND-NAUER R., IIDA S.;
 RT "sequence of the site-specific recombinase gene cin and of its
 RT substrates serving in the inversion of the C segment of bacteriophage
 RT P1";
 RL EMBO J. 2:1733-1740(1983).
 DR EMBL; AJ000741; CAA04283.1; -;
 SQ SEQUENCE 639 AA; 69479 MW; 098D0B57 CRC32;

Query Match 51.9%; Score 42; DB 9; Length 639;
 Best Local Similarity 53.3%; Pred. No. 81;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 AEXNKXIEINTXLD 18
 DB 568 AEANKAIEYLNVMND 582

RESULT 11
 ID 051478 PRELIMINARY; PRT; 315 AA.
 AC 051478;
 DT 01-JUN-1998 (TRENBLrel. 06, Created)
 DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
 DE ALDOSE REDUCTASE, PUTATIVE.
 GN BB0528.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE; 98065943.
 RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
 RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
 RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
 RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
 RA VAN VUOT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
 RA UTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
 RA SMITH H.O., VENTER J.C.;
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
 RT burgdorferi";
 RL Nature 390:580-586(1997).
 DR EMBL; AF001154; AAC6881.1; -;
 DR TIGR; BB0528;
 DR PFAM; PF00246; aldo_ket_red; 2. E17B9A70 CRC32;
 SQ SEQUENCE 315 AA; 36754 MW; E17B9A70 CRC32;

Query Match 50.6%; Score 41; DB 2; Length 198;
 Best Local Similarity 44.4%; Pred. No. 39;
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LKAEKXKXIEINTXLDN 19
 DB 119 LKNEFNAGIDGLNTQIEN 136

RESULT 13
 ID 013950 PRELIMINARY; PRT; 227 AA.
 AC 013950;
 DT 01-JUN-1998 (TRENBLrel. 06, Created)
 DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
 DT 01-JUN-1998 (TRENBLrel. 06, Last annotation update)
 DE PUTATIVE SIGNAL RECOGNITION PARTICLE RECEPTOR BETA SUBUNIT (SR-BETA).
 GN SPAC23H4.07C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA BROWN D., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HAS GTPASE ACTIVITY.
 CC MAY MEDIATES THE MEMBRANE ASSOCIATION OF SR ALPHA
 CC (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM MEMBRANE (POTENTIAL).
 CC -1- SIMILARITY: TO OTHER SRP BETA SUBUNITS; DISTANTLY RELATED TO RAS
 CC SUPERFAMILY.
 DR EMBL; Z98977; CAB11661.1; -;
 KW Hypothetical protein; Signal recognition particle; Transmembrane;
 FT Receptor; Endoplasmic reticulum; GTP-binding.
 FT TRANSMEM 8 28 POTENTIAL.
 FT NP_BIND 44 51 GTP (POTENTIAL).

FT NP_BIND 84 88 GTP (POTENTIAL).
SQ SEQUENCE 227 AA; 25612 MW; 04D2F908 CRC32;

Query Match 50.6%; Score 41; DB 3; Length 227;
Best Local Similarity 47.4%; Pred. No. 45;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 LLKAEKNKXIEEINTXLDN 19
| | | | : | | | : | | :
Db 164 LLKAEKNKXIEEINTXLDN 182

RESULT 14

Q48956 PRELIMINARY; PRT; 87 AA.
AC Q48956;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE 50S RIBOSOMAL PROTEIN (FRAGMENT).
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Capricolum group.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 27343(KID);
RX MEDLINE; 96059641.
RA BORK P., OUZOUNIS C., CASARI G., SCHNEIDER R., SANDER C., DOLAN M.,
RA GILBERT W., GILLET P.M.;
RT "Exploring the Mycoplasma capricolum genome: a minimal cell reveals
its physiology."
RL Mol. Microbiol. 16:955-967(1995).
DR EMBL; 233011; CAA83694.1; -
DR PFAM; PF00831; Ribosomal_L29; 1.
DR PROSITE; PS00579; RIBOSOMAL_L29; 1.
KW Ribosomal protein.
FT NON_TER 87
SQ SEQUENCE 87 AA; 9979 MW; C9A939FD CRC32;

Query Match 49.4%; Score 40; DB 2; Length 87;
Best Local Similarity 46.2%; Pred. No. 26;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 LLKAEKNKXIEE 13
: | | : | | : | | :
Db 68 VIKADYNKAVEEV 80

RESULT 15

Q92610 PRELIMINARY; PRT; 216 AA.
AC Q92610;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE VARIABLE SMALL PROTEIN B.
GN VSPB.
OS Borrelia turicatae.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OZ1;
RA PENNINGTON P.M., CADAVID D., BUNIKIS K., NORRIS S., BARBOUR A.G.;
RT "Interplasmidic arm duplications determine virulence phenotype of the
bacterium Borrelia turicatae."
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049852; AAD15782.1; -
SQ SEQUENCE 216 AA; 22517 MW; 148A1F41 CRC32;

Query Match 49.4%; Score 40; DB 2; Length 216;

Best Local Similarity 43.8%; Pred. No. 60;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 KAEKNKXIEEINTXLD 18
| : | : | | : | : | : | :
Db 171 KTRGAKLELELTAID 186

Search completed: November 13, 1999, 12:55:44
Job time: 3043 sec

THIS PAGE BLANK (USPTO)